

2025 年度

修 士 論 文

**感情予測モデルを用いた  
日英間の絵文字使用の文化差分析**

指導教員：村上 陽平

立命館大学大学院 情報理工学研究科  
情報理工学専攻 博士課程前期課程  
計算機科学コース

学生証番号：6611240060-6

氏名：森 叶葉

## 感情予測モデルを用いた日英間の絵文字使用の文化差分析

森 叶葉

### 内容梗概

絵文字をテキストに組み込むことで、テキストに書き手の感情やイメージを付与することができる。機械翻訳を用いた多言語コミュニケーションでは、メッセージが翻訳され表現が変わるため、絵文字は書き手の感情や意図を直接伝える重要な役割を果たす。しかしながら、絵文字は同じ文字であっても文化によって異なる解釈をされる場合がある。特に、機械翻訳において絵文字はそのまま挿入されるため、異言語間で誤解を生じさせやすい。

そこで本研究では、日本語と英語における絵文字に対する感情の文化的な違いを定量的に抽出する手法を提案する。具体的には、 $X$  から収集した絵文字が付与されたテキストを感情予測モデルに入力し、8つの感情ごとに予測値を得る。その後、絵文字ごとに予測値の平均を取ることで、絵文字が抱く感情を8次元のベクトルで表現する。本研究では、このベクトルを絵文字感情ベクトルと呼ぶ。日英の絵文字感情ベクトル間の違いからの絵文字の文化差を抽出する。手法の実現にあたり、取り組むべき課題は以下の2点である。

### 絵文字感情予測モデルの構築

絵文字感情を定量化するためには、絵文字感情をテキストから抽出する必要がある。表情や感嘆疑問符は感情と関連付けやすいが、動物や植物などの絵文字は感情と関連付けにくい。そこで、絵文字が付与されているテキストの本文を活用し、どのような絵文字に対しても感情の定量化を行うモデルが必要である。

### 絵文字感情の文化差抽出方法

絵文字感情の文化差の評価を人手で行うと、個人の主観や文化知識が影響するため、評価に揺らぎが生じる。そのため、人手による評価を用いず統計的手法により文化差を抽出する必要がある。また、統計的手法により抽出された文化差のある絵文字が実世界での解釈と整合するかを確かめるため、人間の判断を参照し、文化差抽出結果の妥当性を評価する必要がある。

1つ目の課題に対しては、入力文に込められた感情を予測する感情予測モデルを構築し、絵文字が付与されたテキストの感情を予測することでその絵文字感情の分散表現を獲得した。具体的には、複数の感情の有無に対して予測ができるよう、Plutchikの8種類の感情カテゴリ(喜び, 悲しみ, 期待, 驚き, 怒

り、恐れ、嫌悪、信頼)が付与された主観と客観の感情分析データセットの主観のデータを用いて、事前学習された BERT のファインチューニングを行った。次に、X から絵文字を含む日本語と英語のポストを収集し、感情予測モデルを用いて絵文字付きテキストの感情を予測し、その絵文字の絵文字感情ベクトルを構築する。

2 つ目の課題に対しては、絵文字の文化差を抽出するために、抽出観点の異なる 3 つの分析手法を用いた。まず、絵文字感情ベクトルを構成する 8 つの感情軸それぞれに着目し、各感情の強さの量的な違いに基づく文化差を t 検定により抽出した。次に、単一感情軸では捉えきれない感情の方向性の違いを補完するため、対極に位置する感情ペアの差に着目し、その差が日英間で異なるかを t 検定により検証した。さらに、これらの手法でも抽出しきれない感情構造全体の違いを捉えるため、8 次元の感情の順位構造に基づいてスピアマンの順位相関係数を算出し、文化差を抽出した。それぞれの文化差を順に、局所的文化差(第一種文化差)、対極的文化差(第二種文化差)、構造的文化差(第三種文化差)と呼ぶ。加えて、人手による絵文字感情アノテーション結果を使用し、統計的に抽出された文化差の妥当性を確認した。

本研究の貢献は以下の通りである。

### 絵文字感情モデルの構築

構築した感情予測モデルは、テキストに付与された感情を 8 感情のカテゴリから予測する。感情ごとの予測値は最適閾値を使用した。日本語モデルの適合率、再現率、F 値のマクロ平均は 0.47, 0.70, 0.56 であった。英語モデルの適合率、再現率、F 値のマクロ平均は 0.48, 0.66, 0.56 であった。

### 絵文字感情の文化差抽出手法

絵文字感情ベクトルを用いた文化差抽出の結果、3 つの観点から日英間で文化差のある絵文字を 305 件抽出した。さらに、統計的に抽出された文化差のある絵文字の妥当性を検証するため、人手アノテーションに基づく絵文字感情ベクトルから文化差のある絵文字を抽出し、提案手法による結果との一致度を評価した。その結果、人手ベースの文化差のある絵文字は 431 件であり、両方で文化差ありの判断が一致した絵文字は 137 件であった。このとき適合率は 0.45、再現率は 0.32、F 値は 0.37 となり、提案手法が一定数の文化差のある絵文字を自動的に抽出できる一方で、人が捉える文化差を十分に網羅できていない課題も示された。

## **Analysis of Cultural Difference in Emoji Usage Between Japanese and English Based on Emotion Prediction Models**

Kanaha Mori

### **Abstract**

By inserting emojis into text, writers can convey emotions and imagery through their writing. Emojis play a crucial role in conveying emotions and intentions in multilingual communication, particularly in contexts where machine translation (MT) may change. However, emojis may be interpreted differently depending on the culture, even when using the same emoji. Thus, emojis are likely to cause misunderstandings between languages, since they are inserted directly into MT.

Therefore, we propose a method to quantitatively extract cultural differences in emotion attributed to emoji in Japanese and English. Specifically, text containing emojis collected from X is input into an emotion prediction model to obtain predicted values for each of the eight emotions. Next, by averaging these values for each emoji, the emotion associated with the emoji is represented as an 8-dimensional vector. These vectors are called the emoji emotion vector in our research. To achieve this method, we address the following two challenges.

### **Constructing an Emotion Prediction Model**

To quantify emoji emotions, it is necessary to extract them from text. However, for emojis such as animals and plants, it is difficult to link them to emotions compared to facial emoji and exclamation or question marks. Therefore, a model is needed that utilizes the context of the text containing emoji to quantify emotions for any type of emoji.

### **Method for Detecting Cultural Differences in Emoji Emotion**

Manual evaluation of cultural differences in emoji emotions causes variation in assessment due to individual subjectivity and cultural knowledge. Therefore, it is necessary to extract cultural differences using statistical methods rather than manual evaluation. Furthermore, to verify that emojis extracted by statistical methods match real interpretations, we assess the validity of these results by referencing human judgments and actual usage contexts.

For the first challenge, an emotion prediction model was built to obtain distributed representations of emoji emotions. A pretrained BERT model was

fine-tuned on a subjective emotion dataset annotated with Plutchik’s eight categories. Japanese and English emoji-containing posts were then collected from X, and emoji emotion vectors were constructed by predicting emotions in these texts.

For the second challenge, cultural differences in emoji usage were extracted using three complementary methods. First, t-tests were applied to each emotion dimension to detect quantitative differences in emotional intensity. Second, differences between opposite emotion pairs were examined to capture directional variations. Finally, Spearman’s rank correlation was used to identify structural differences in overall emotion rankings. The validity of these results was further confirmed through human-annotated emoji emotion data.

The contributions of this study are summarized as follows.

### **Constructing an Emotion Prediction Model**

The constructed emotion prediction model predicts emotions assigned to a text from eight emotion categories. For each emotion, the predicted scores were binarized using an optimal threshold. The Japanese model achieved macro-averaged precision, recall, and F1-score of 0.47, 0.70, and 0.56, respectively. The English model achieved macro-averaged precision, recall, and F1-score of 0.48, 0.66, and 0.56, respectively.

### **Method for Detecting Cultural Differences in Emoji Emotion**

As a result of cultural difference extraction, 305 emojis exhibiting cultural differences between Japanese and English were identified from three analytical perspectives. Furthermore, to validate the statistically extracted culturally different emojis, culturally different emojis were also extracted from human-annotated emoji emotion vectors, and the agreement with the proposed method was evaluated. As a result, 431 emojis were identified as culturally different in the human-based analysis, and 137 emojis were commonly judged as culturally different by both approaches. In this evaluation, the precision was 0.45, the recall was 0.32, and the F1-score was 0.37. These results indicate that the proposed method can automatically extract a certain number of culturally different emojis, while also revealing a limitation in that it does not sufficiently cover the cultural differences perceived by humans.

# 感情予測モデルを用いた日英間の絵文字の文化差分析

## 目次

<b>第1章 はじめに</b>	<b>1</b>
<b>第2章 絵文字コミュニケーション</b>	<b>3</b>
2.1 絵文字の役割	3
2.2 絵文字の違い	3
2.3 本研究の位置付け	4
<b>第3章 感情予測モデルの構築</b>	<b>6</b>
3.1 BERT	6
3.1.1 事前学習モデル	6
3.1.2 ファインチューニング	7
3.2 モデルの構築	7
3.2.1 感情ラベル付きデータセット	8
3.2.2 最適エポック数の算出	8
3.2.3 最適閾値の算出	13
3.3 感情予測モデルの精度評価	14
3.3.1 弱感情データから作成した感情予測モデルの精度	14
3.3.2 中感情データから作成した感情予測モデルの精度	16
3.3.3 強感情データから作成した感情予測モデルの精度	17
<b>第4章 絵文字感情の分散表現</b>	<b>19</b>
4.1 絵文字付きデータセット	19
4.2 感情予測に基づく絵文字感情分散表現	19
<b>第5章 絵文字の文化差抽出</b>	<b>21</b>
5.1 感情軸に合わせたレーダーチャートによる視覚化	21
5.2 局所的文化差の抽出	22
5.3 対極的文化差の抽出	22
5.4 構造的文化差の抽出	26
<b>第6章 評価</b>	<b>29</b>
6.1 人手での絵文字感情アノテーション	29

6.2 絵文字感情分散表現の正当性.....	34
6.3 人手による文化差のある絵文字の定義 .....	35
6.4 提案手法によって抽出された文化差の妥当性 .....	39
<b>第7章 考察</b>	<b>41</b>
7.1 感情予測モデルの特徴と限界 .....	41
7.2 絵文字感情ベクトルの特徴.....	42
7.2.1 感情予測に基づくモデルベースの絵文字感情ベクトル .....	42
7.2.2 人手アノテーションによる人手ベースの絵文字感情ベクトル...	43
7.2.3 両手法による絵文字感情ベクトルの違い .....	43
7.3 提案手法による文化差抽出結果の妥当性.....	44
7.3.1 文化差種類別内訳から見た文化差抽出結果 .....	45
7.3.2 感情別内訳から見た文化差抽出結果.....	47
7.3.3 絵文字カテゴリ別に見た文化差抽出結果 .....	47
7.3.4 分析手法の制約 .....	50
7.4 課題.....	51
<b>第8章 おわりに</b>	<b>53</b>
<b>謝辞</b>	<b>55</b>
<b>参考文献</b>	<b>56</b>
<b>付録</b>	<b>58</b>
A.1 提案手法と人手評価における TP・FP・FN の絵文字一覧 .....	58

## 第1章 はじめに

機械翻訳を介した多言語コミュニケーションにおいて、絵文字は重要な役割を果たす。絵文字はテキストに書き手の感情や意図を付与するため、テキストの読み手は書き手の意図を容易に理解できる。チャットコミュニケーションにおける絵文字の使用は、書き手と読み手の意思疎通の手助けをする効果がある。

しかしながら、機械翻訳を介した多言語コミュニケーションにおいて、絵文字は翻訳されずにそのまま挿入されるため、絵文字の解釈が異なる翻訳先言語では、書き手の意図とは異なる絵文字の解釈をされる可能性がある。例として、の日本語と英語の解釈の違いを挙げる。日本語では鼻提灯を垂らしながら眠っている顔として解釈されることが多いが、英語では鼻提灯の部分を涙と解釈され、悲しみを表すときにこの絵文字を用いることがある。日本語話者が「ぐっすり眠れた😓」と入力すると、英語への翻訳文は「I could slept soundly😓」となる。日本語話者はぐっすり眠れたことによる喜びを示すために「😓」を入力するが、英語話者が「😓」が泣いている絵文字であると解釈した場合、悲しんでいる表現の絵文字「😓」が、喜んでいる表現のテキスト「I could slept soundly」と結び付かず、英語話者は日本語話者のメッセージの意図を理解することが困難になる。このように、文化圏によって解釈の異なる絵文字をそのまま挿入すると、読み手が書き手の感情や意図を正しく理解できず、コミュニケーション齟齬の原因となる。したがって、多言語コミュニケーションにおいて、テキストだけでなく絵文字も翻訳される必要がある。絵文字翻訳の実現には、異言語間での絵文字解釈の文化差を抽出する必要がある。

そこで本研究では絵文字で表現される感情を日本語と英語 2 つの言語ごとに個々のベクトルで表現し、その値に基づいて日英間のコミュニケーション齟齬に繋がる絵文字の解釈の文化差を抽出する。具体的には、複数の感情の有無に対して予測ができるよう、8 種類の感情カテゴリが付与された主観と客観の感情分析データセットの主観のデータを用いて、事前学習された BERT のファインチューニングを行った。次に、X から絵文字を含む日本語と英語のポストを収集し、感情予測モデルを用いて絵文字付きテキストの感情を予測し、その絵文字の絵文字感情ベクトルを構築する。続いて、構築した日英の絵文字感情ベクトルを使用し、3 種類の統計的な分析手法により文化差の抽出を行った。また、統計的に文化差が抽出できた絵文字に対して、日本語話者と英語話者による絵文字感情

によるアノテーションを行い、そのデータを照合することで抽出された文化差の妥当性の評価を行う。本研究において取り組む課題は、絵文字感情を定量的に算出する絵文字感情予測モデルの作成、統計的な絵文字感情の文化差の抽出方法の2点である。本研究では、絵文字解釈の文化差を定量的に抽出するために、主に2つの課題に取り組む。1点目の課題については、絵文字には表情や感嘆疑問符のように感情と直接結びつくものだけでなく、動物・食べ物・物体など感情と結びつきにくい絵文字も存在する。そこで本研究では、絵文字単体ではなく絵文字が付与されているテキストの本文を活用し、テキストに含まれる感情を8種類の感情カテゴリ(喜び, 悲しみ, 期待, 驚き, 怒り, 恐れ, 嫌悪, 信頼)として予測できるモデルを構築する。次に2点目課題については、絵文字解釈の文化差を人手のみで評価する場合、主観や文化的背景により判断が揺らぐ可能性がある。そのため本研究では、感情軸ごとの量的な差, 対極に位置する感情ペアの差, 8つの感情全ての順位構造の違いといった複数の観点から文化差候補を抽出する。さらに、人手アノテーション結果を参照することで、統計的に抽出された文化差が実際の人間の解釈と整合するかを検証する。

以下、本論文では、第2章で絵文字のチャットコミュニケーションにおける役割と絵文字解釈の文化差について述べる。次に、第3章で絵文字感情を埋め込むときに使用する感情予測モデルの構築について述べる。第4章で、第3章にて作成した感情予測モデルを用いた絵文字感情ベクトルの作成方法について述べる。第5章で、第4章で作成した日英の絵文字感情ベクトルから、多面的な絵文字の文化差を統計的に抽出する手法について述べる。第6章では、人手での絵文字に対する感情アノテーション結果を使用して、第5章で統計的に抽出した絵文字の文化差が、人が感じる文化差とどの程度一致しているのかを確認し、第7章で考察を行う。最後に、第8章で今後の展望について述べる。

## 第2章 絵文字コミュニケーション

### 2.1 絵文字の役割

絵文字は、コミュニケーションの中で重要な役割を果たしている。Thomasらは、デジタルコミュニケーションで絵文字が使用されることで、話し手の間接的な意味の認識を促進する可能性について研究を行っている[1]。一部の絵文字について、受け手に間接的な意味をより頻繁に、迅速に認識させる促進効果があることを示している。テキストと絵文字の両方を含む状況では、絵文字が対面での表情のように受け手に間接的な意味を認識させる役割を果たす可能性が高いことを示唆している。

さらに、絵文字と感情に着目した研究も行われている。Petraらは、絵文字感情辞書を作成し、絵文字が持つ感情をポジティブ、ネガティブ、ニュートラルのスコアとして表現している[2]。また、この研究では絵文字を含むツイートと含まないツイートの人間の知覚の相互作用の分析も行われており、絵文字の存在は人間のツイートの感情的な知覚に良い影響を与えると結論づけられている。メッセージの送信者、受信者双方の視点から、MMCにおける絵文字使用について研究を行っている。送信者は一般的に絵文字の使用は、感情の表現、表現の強化、テキストの文脈のトーンの調整を意図していることが示されている。

### 2.2 絵文字の違い

絵文字使用や絵文字解釈の違いに関する研究も行われてきた。Zhaoらは、絵文字の使用状況、ツイートの内容、構造、ユーザの人口統計の相関係数を調査し、そこで得られた観察をもとに絵文字とツイート内容、ツイート構造、ユーザー人口統計学の間を相関を定量的に記述するためのmmGRUを開発し、そこから予測された結果を分析している[3]。その結果、地域によって絵文字の選択に違いがあること、同じ絵文字でも地域によって感情の捉え方に違いがあること、絵文字の種類によってテキスト中で使用される位置が異なる傾向があることが報告されている。Wesleyらは、性別と年齢が感情に関連する絵文字の解釈に与える影響の妥当性を明らかにするオンライン調査を行っている[4]。9つの絵文字に対してそれぞれの絵文字はどのような感情を示すかという質問に、参加者に答えるように求め、言語を超えて自由形式の回答が許可されていた。各回答は直接

マッピングしたり、同義語を決定したり、手作業でラベリングを行ったりすることで、各絵文字が示す感情が Plutchik の基本 8 感情にラベリングされた。その結果、性別によって解釈が異なる絵文字(Unasumed Face😞, Nauseated Face🤢, Face with Rolling Eyes🙄)があることが示されている。また、年齢によって Unasumed Face😞の絵文字の解釈は、年齢のカテゴリによって異なることが示唆されている。Vanessa は、実際に絵文字翻訳に取り組んだ例から絵文字の違いについて、Semenova による A Christmas Carol の絵文字翻訳の抜粋を取り挙げ分析している[5]。同じ絵文字でも使用されるテキストの組み合わせによってその絵文字の意味が全く異なり、普遍的な理解が難しいことを示している。Cho らは、文化の違いによるピクトグラムの解釈について、日本とアメリカでオンライン調査を行っている[6]。その結果、19 種類のピクトグラムに対して日米で文化的に異なる解釈を持つことが報告されている。これらの解釈の違いの分析によると、解釈の違いは以下の 3 つのパターンに分類されると示されている。

- 2 つの文化は同じ基礎概念を共有しているが、概念について異なる視点を持っている。
- 2 つの文化は同じ基礎概念を部分的にしか共有していない。
- 2 つの文化は基礎概念を共有していない。

### 2.3 本研究の位置付け

このようにこれまで絵文字使用に関する文化間の違いについて分析されているが、使用頻度や好みの傾向、受け手の解釈における違いなどが主に調査されてきた。しかし、絵文字を用いた日英間の感情表現の違いについて、複数の感情を同時に扱いながら定量的に比較した研究は十分に行われていない。特に、既存研究の多くは、単一の感情ラベルやポジティブ・ネガティブ・ニュートラルといった分類に基づく分析に留まっており、絵文字が文脈に応じて複数の感情を同時に表現する点については十分に扱われてこなかった。また、人手アノテーションに基づく分析では、評価者の主観や属性の影響を受けやすく、大規模な使用実態を反映することが難しいという課題も存在する。

そこで本研究では、実際の SNS 投稿を用いてマルチラベル分類を行う感情予測モデルを構築し、投稿文脈に基づいて絵文字に付随する感情傾向を多次元ベクトルとして表現する手法を提案する。さらに、提案手法によって得られた絵文字感情ベクトルを、人手アノテーションによる結果と比較することで、日英間に

における絵文字の感情表現の違いを定量的に分析する。本研究の新規性は、マルチラベル感情予測に基づいて絵文字の感情を多次元的に捉える点、実使用データに基づく大規模な分析を行う点、人手アノテーションとの比較を通じて結果の解釈を補強する点にある。これにより、従来の分析や主観的評価に留まらない、絵文字を用いた感情表現における文化差の理解を目指す。

## 第3章 感情予測モデルの構築

Dresner らによると、日常的なデジタル媒介の会話で絵文字は、文字による言語コミュニケーションに組み込まれており、相互関係を持つ[7]。そこで、私たちは絵文字の文化差を抽出するために、「ある絵文字付きテキストに込められた感情は、その絵文字に関連がある」と仮説を立て、絵文字に付与された感情を定量的に表現することで、文化差の抽出を行う。具体的には、絵文字感情を定量化するために感情予測モデルを、自然言語処理モデル **BERT** を用いて構築する。本章では、絵文字感情の定量化に使用する **BERT** の概要、本研究で構築した感情予測モデルの構築方法、構築したモデルの精度評価について述べる。

### 3.1 BERT

**BERT**[8]とは、2018年にGoogleによって発表された自然言語処理の手法であり、多層のTransformerエンコーダブロックで構成されている。従来の自然言語処理モデルのRNNやLSTMと違い、**BERT**はTransformerに基づく自己注意機構を用いることで、文中の単語間の関係を効率的に捉え、文脈に基づいた表現を生成することができる。

#### 3.1.1 事前学習モデル

**BERT**は事前学習とファインチューニングの2段階の学習を行うことが大きな特徴である。モデルを1から学習させることは、膨大な時間とコストがかかる。そこで**BERT**では、言語の広範囲な知識を身につけた事前学習モデルを利用して、求めるタスクに特化したモデルを作成することが一般的である。

事前学習では、大規模なデータコーパスを用いてMasked Language Model(MLM)とNext Sentence Prediction(NSP)の2種類のタスクが用いられることが多い。MLMは、単語の双方向での文脈理解を目的に行うタスクである。具体的には、入力文の一部の単語を[MASK]トークンに置き換え、この[MASK]トークンを予測する学習を行う。NSPは、連続する2文の関連性を理解することを目的に行うタスクである。具体的には、2つの文をペアとして入力し、その2文が連続しているかどうかを判定するタスクを用いて学習を行う。これらの事前学習によって、単語や文脈をより考慮できるモデルを獲得している。そのため、従来の自然言語処理モデルと比べて**BERT**は、事前学習を用いたファインチューニングによって、低コストでの転移学習が可能となっている。

### 3.1.2 ファインチューニング

ファインチューニングでは、個別のタスクのラベルが付いたデータから、BERT がそのタスクに特化するように学習を行う。本研究では、テキストから Plutchik の基本 8 感情を複数予測するタスクを実現するため、8 感情それぞれについて独立に予測を行うマルチラベル分類として定式化する。マルチラベル分類とは、文章が与えられたとき、選択肢の中から複数のカテゴリを選択する分類のことである。本研究では、テキストから Plutchik の基本 8 感情を複数予測するタスクを実現するため、8 感情それぞれについて独立に予測を行うマルチラベル分類として定式化する。具体的な本研究でのファインチューニングでは、テキストから 8 感情にマルチラベル分類を行うタスクを学習させる。

### 3.2 モデルの構築

日本語のモデル構築と英語のモデル構築に使用する BERT の事前学習モデルは、それぞれ、東北大学の乾研究室が公開している日本語 BERT モデル<sup>i</sup>、Google が公開している英語 BERT モデル<sup>ii</sup>を使用する。BERT には、文脈をより考慮できる特徴があり、テキストの文脈をもとに感情を予測するタスクを行う感情予測モデルを構築するには最適である。事前学習モデルを、テキストから感情ラベルにマルチラベル分類ができるようにファインチューニングした。具体的には、特徴量をテキスト、ラベルを Plutchik の基本 8 感情とする。また、出力が各感

表 1 弱感情データの一部

sentence	ぼけっとしてたらこんな時間。 チャリあるから食べにでたいのに…	
感情ラベル	元データ	置換後
喜び	0	0
悲しみ	1	1
期待	2	1
驚き	1	1
怒り	1	1
恐れ	0	0
嫌悪	0	0
信頼	1	1

<sup>i</sup> cl-tohoku/bert-base-japanese

<sup>ii</sup> Bert-base-uncased

情ラベルとして表されるように設定した。テキストに含まれる単語や文脈の情報をもとに、感情を予測するタスクを繰り返すことで、モデルはテキストと感情の関連を学習する。学習データには、感情ラベルが付与された感情ラベル付きデータ 34560 件を用いる。感情強度を二値化した 3 種類のデータセットについて感情予測モデルを作成し、最も精度が良いモデルを 4 章の絵文字感情ベクトルの作成に用いる。

### 3.2.1 感情ラベル付きデータセット

感情予測モデルを構築するために必要な、感情ラベルがテキストに付与されたデータセットを作成する。梶原らが公開している、テキストに感情強度をラベル付けした主観と客観の感情分析データセット[9]を用いる。主観と客観の感情分析データセットとは、テキストの書き手の主観的な感情強度とテキストの読み手の客観的な感情強度のラベルを収集したデータセットである。ここでの感情は、Plutchik の基本 8 感情(喜び, 悲しみ, 驚き, 怒り, 恐れ, 嫌悪, 期待, 信頼) を指す[10]。本論文では、多言語チャットコミュニケーションにおいて、テキストの書き手がテキストに込める感情に着目するため、主観と客観の感情分析データセットのうち、主観の感情分析データセットを用いる。主観の感情分析データは、テキストの書き手が 8 種類の感情を段階の強度でそれぞれラベル付けを行っている。英語の感情ラベル付きデータセットを Google 翻訳で機械翻訳して作成した。モデルのファインチューニングに用いる際に、感情のラベル付けを二値にする必要がある。そこで、付与された感情強度を二値に置換して使用した。以降、感情強度が 0 のものを「0」、1 以上のものを「1」と置換したデータを弱感情データ、感情強度が 0, 1 のものを「0」、2 以上のものを「1」と置換したデータを中感情データ、感情強度が 0, 1, 2 のものを「0」、3 のものを「1」と置換したデータを強感情データと呼ぶ。表 1 は、弱感情データの一文を抜粋したものである。元データの値から、感情強度が 0 のものを「0」、1 以上のものを「1」と置換している。ここで作成した 3 種類のデータセットを使用し、感情予測モデルを構築する。

### 3.2.2 最適エポック数の算出

最適なエポック数は、学習の設定は、バッチサイズ 64 とし、バリデーションロスが下がり、かつ、暫定の閾値を 0.5 とした時の F 値が最も高いエポック数を採用した。

図 1 から図 6 で、3 種類それぞれのデータセットを用いて日英の感情予測モ

デルを構築したとき、エポック数ごとの損失と F 値についてのグラフを示す。図 1 を見ると、エポック数が増加するにつれトレーニングロスが減っているものの、エポック 3 からはバリデーションロスが増加しており、F 値も滑らかながら減少傾向にある。そのため、弱感情データを用いた日本語の感情予測モデルの最適なエポック数は 2 と仮定する。図 2 を見ると、図 1 の場合と同様にエポック数が増加するにつれトレーニングロスが減っているものの、エポック 3 からはバリデーションロスが増加しており、F 値も減少している。そのため、弱感情データを用いた英語の感情予測モデルの最適なエポック数も 2 と仮定する。続いて、中感情データを用いて作成した感情予測モデルでは、図 3 より、日本語ではトレーニングロスが増加させるにつれ減少しているが、バリデーションロスはエポック数が 3 になると増加しており、F 値も減少している。そのため、中感情データを用いた日本語の感情予測モデルの最適なエポック数は 2 と仮定する。図 4 の英語の感情予測モデルでは、トレーニングロスが増加させるにつれ減少しているが、バリデーションロスはエポック数が 2 になると増加しており、F 値も減少している。そのため、中感情データを用いた英語の感情予測モデルの最適なエポック数は 1 と仮定する。強感情データを用いて作成した図 5 の日本語モデルでは、トレーニングロスが増加させるにつれ減少しているが、バリデーションロスはエポック数が 2 になると増加しており、F 値も減少している。そのため、このモデルの最適なエポック数は 1 と仮定する。最後に、図 6 の英語モデルでは、トレーニングロスが増加させるにつれ減少しているが、バリデーションロスはエポック数が 3 になると増加しており、F 値も減少している。そのため、このモデルの最適なエポック数は 2 と仮定する。ここで仮定したデータセットごとの最適エポック数を使用して、3.2.3 章で感情ラベルごとの最適閾値を算出する。

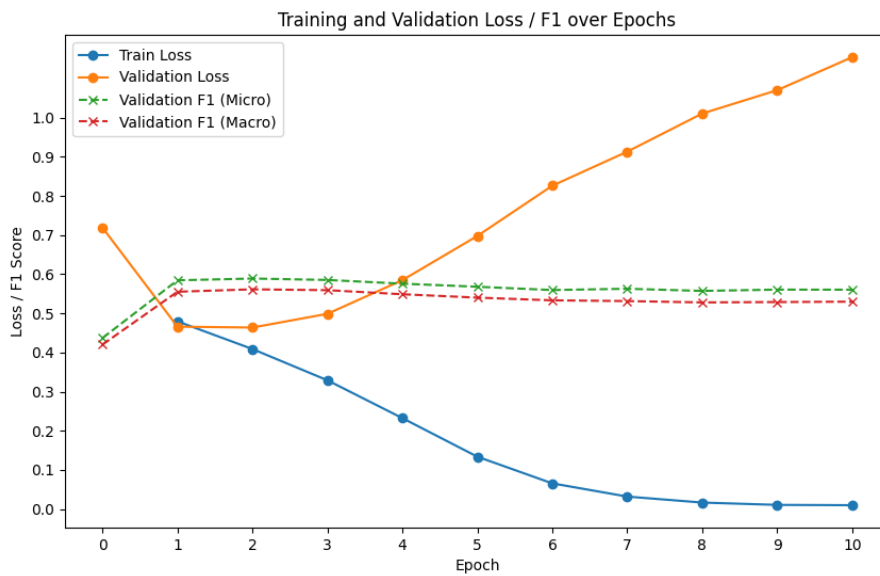


図 1 弱感情データにおける日本語の感情予測モデルに対するエポック数ごとの loss と F 値

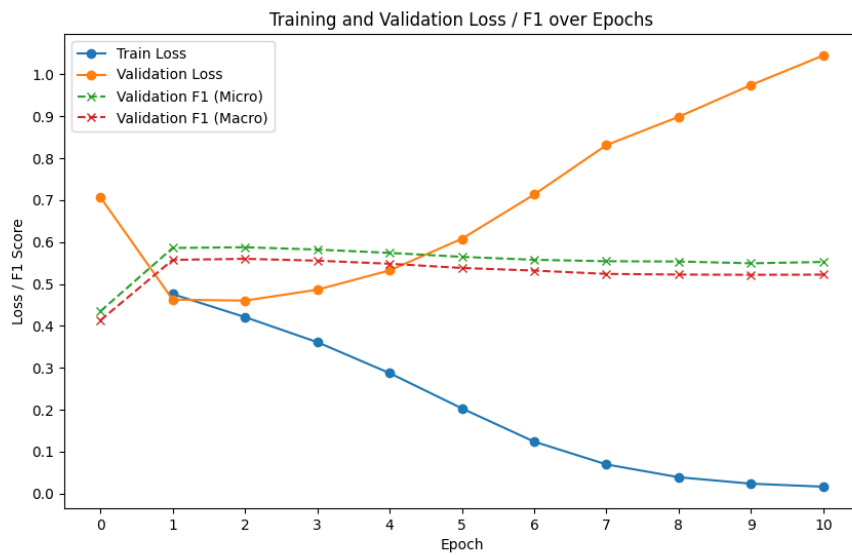


図 2 弱感情データにおける英語の感情予測モデルに対するエポック数ごとの loss と F 値

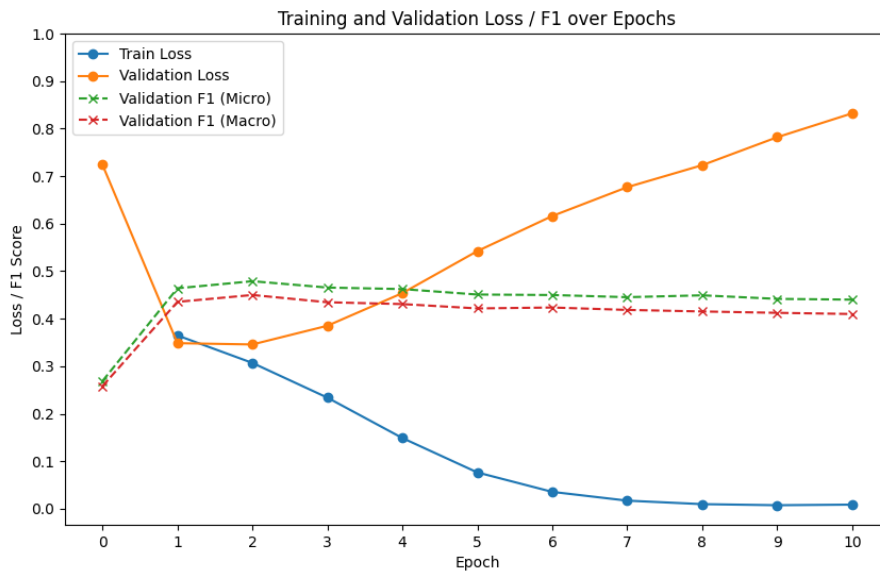


図 3 中感情データにおける日本語の感情予測モデルに対するエポック数ごとの loss と F 値

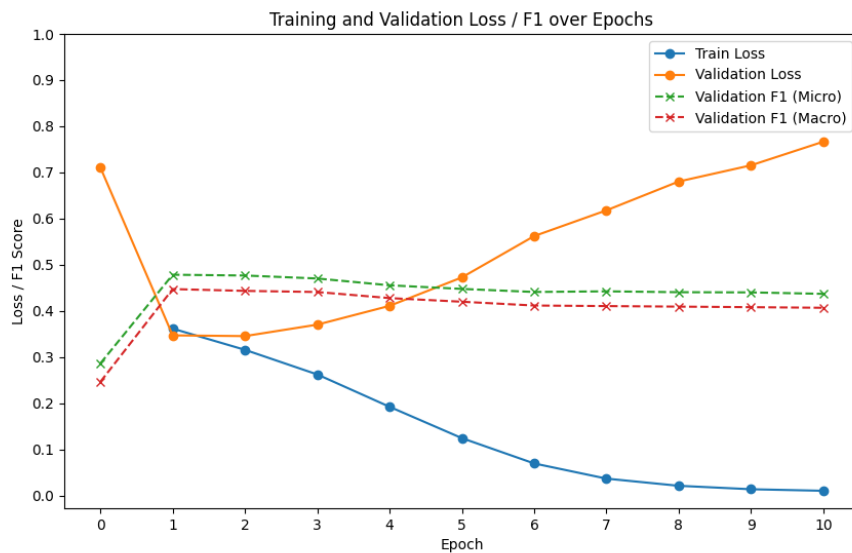


図 4 中感情データにおける英語の感情予測モデルに対するエポック数ごとの loss と F 値

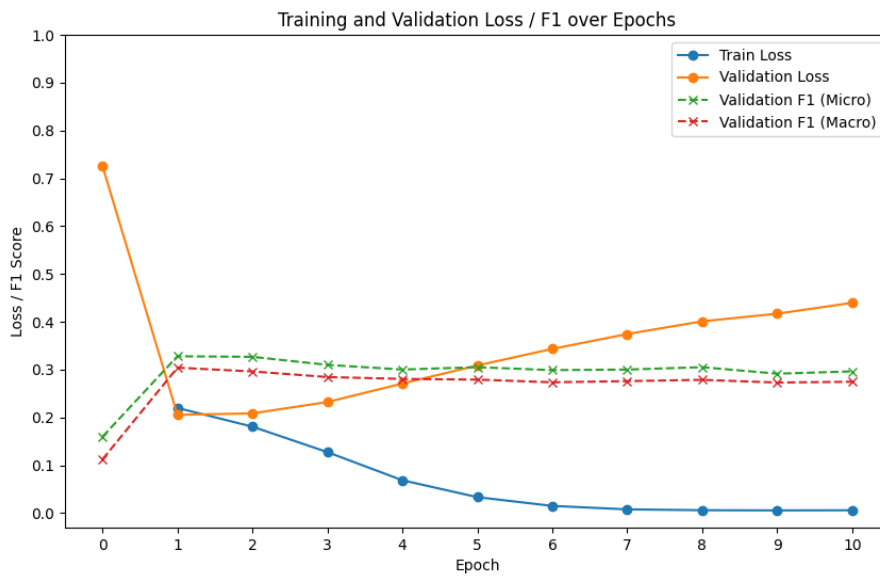


図 5 強感情データにおける日本語の感情予測モデルに対するエポック数ごとの loss と F 値

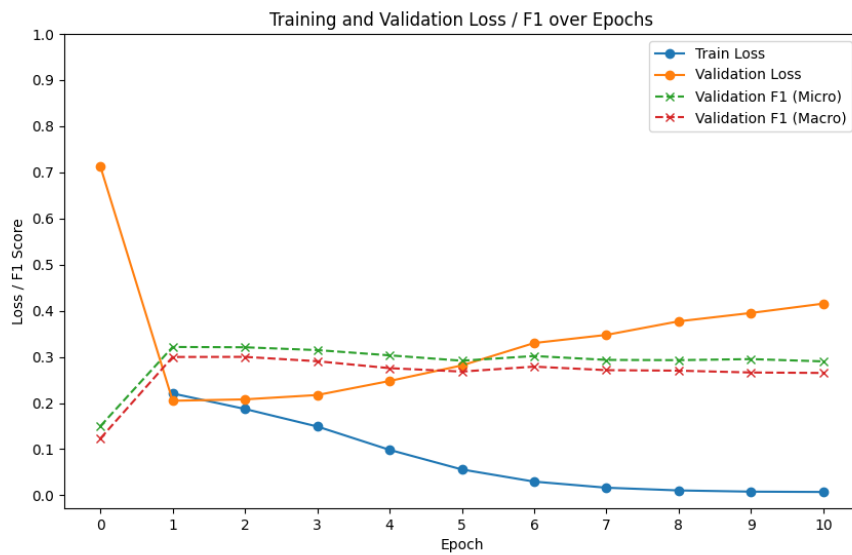


図 6 強感情データにおける英語の感情予測モデルに対するエポック数ごとの loss と F 値

### 3.2.3 最適閾値の算出

3.2.2 章で算出した、データセット、言語ごとに最適なエポック数を設定し作成したモデルについて、検証データから F 値を算出した。具体的には、各感情レベルの予測確率に対して閾値を変化させながら適合率、再現率を算出し、F 値が最大となる閾値をその感情の最適閾値として採用した。感情の予測確率については、それぞれの最適閾値以上の感情は「1」、それ未満の感情は「0」と設定した。

表 2 から表 4 にデータセット別に日英のモデルで感情ごとの最適閾値を算出した結果を示す。表 2 を見ると、日本語の最適閾値の方が英語の最適閾値に比べ比較的高いが、言語間で最適閾値に大きな差はない。また、感情ごとに最適閾値を確認すると、日本語では悲しみの閾値が 0.407 と他の感情に比べて最も大きく、怒りの閾値が 0.214 で最も小さい。英語では日本語と同様、悲しみの閾値 0.368 で最も大きく、信頼の閾値が 0.198 で最も小さい。次に、表 3 を見ると、日本語の最適閾値の方が英語の最適閾値に比べ比較的高いが、言語間で最適閾値に大きな差はない。また、感情ごとに最適閾値を確認すると、日本語では悲しみの閾値が 0.407 と他の感情に比べて最も大きく、怒りの閾値が 0.214 で最も小さい。英語では日本語と同様、悲しみの閾値 0.368 で最も大きく、信頼の閾値が 0.198 で最も小さい。続いて、表 4 を見ると、日本語の最適閾値の方が英語の最適閾値に比べ比較的高いが、言語間で最適閾値に大きな差はない。また、感情ごとに最適閾値を確認すると、日本語では悲しみの閾値が 0.407 と他の感情に比べて最も大きく、怒りの閾値が 0.214 で最も小さい。英語では日本語と同様、悲しみの閾値 0.368 で最も大きく、信頼の閾値が 0.198 で最も小さい。各データセッ

表 2 弱感情データにおける感情ごとの最適閾値

	日本語	英語
喜び	0.274	0.232
悲しみ	0.407	0.368
期待	0.236	0.250
驚き	0.327	0.274
怒り	0.214	0.219
恐れ	0.277	0.232
嫌悪	0.266	0.264
信頼	0.258	0.198

表 3 中感情データにおける感情ごとの最適閾値

	日本語	英語
喜び	0.385	0.333
悲しみ	0.265	0.297
期待	0.206	0.278
驚き	0.187	0.191
怒り	0.340	0.200
恐れ	0.161	0.158
嫌悪	0.234	0.177
信頼	0.227	0.217

表 4 強感情データにおける感情ごとの最適閾値

	日本語	英語
喜び	0.168	0.314
悲しみ	0.201	0.125
期待	0.220	0.173
驚き	0.0828	0.107
怒り	0.152	0.137
恐れ	0.106	0.0924
嫌悪	0.184	0.154
信頼	0.0869	0.136

ト、言語ごとに 3.2.2 章で算出した最適なエポック数、3.2.3 章で算出した最適閾値を使用し感情予測モデルを作成する。

### 3.3 感情予測モデルの精度評価

感情予測モデルの精度評価には、4320 件のテストデータを使用した。感情の閾値はデータセット、言語ごとに表 2 から表 4 に示す最適閾値を使用する。感情ごとに適合率、再現率、F 値を算出しそれらのマクロ平均をモデル全体の性能指標とした。

#### 3.3.1 弱感情データから作成した感情予測モデルの精度

表 5 に日本語の感情予測モデルの感情ごとの精度を、表 6 に英語の精度を示す。表 5 のように、日本語のモデルの感情ごとの精度は、全ての感情に対して、適合率よりも再現率の方が高かった。感情ごとに精度を見ると、最も適合率が高い感情は喜びで 0.69、最も再現率が高い感情は悲しみで 0.78、最も F 値が高い

感情は悲しみで 0.71 であった。また、最も適合率が低い感情は怒りで 0.35、最も再現率が低い感情は恐れで 0.53、最も F 値が低い感情は怒りで 0.43 であった。表 6 により、英語モデルの感情ごとの精度についても、日本語モデルと同様に、適合率よりも再現率の方が高かった。感情ごとに精度を見ると、最も適合率が高い感情は喜びで 0.65、最も再現率が高い感情も喜びで 0.82、最も F 値が高い感情は悲しみで 0.73 であった。また、最も適合率が低い感情は信頼で 0.36、最も再現率が低い感情は怒りで 0.53、最も F 値が低い感情は怒りで 0.45 であった。弱感情データから作成した日本語のモデル全体のマクロ平均適合率は 0.49、マクロ平均再現率は 0.67、マクロ平均 F 値は 0.56 であり、英語のモデルのマクロ平均適合率は 0.47、マクロ平均再現率は 0.71、マクロ平均 F 値は 0.56 であった。

表 5 弱感情データから作成した日本語の感情予測モデルの精度

感情	適合率	再現率	F 値
喜び	0.69	0.74	0.71
悲しみ	0.58	0.78	0.67
期待	0.55	0.77	0.64
驚き	0.50	0.66	0.57
怒り	0.35	0.57	0.43
恐れ	0.40	0.53	0.45
嫌悪	0.39	0.75	0.51
信頼	0.45	0.54	0.49
マクロ平均	0.49	0.67	0.56

表 6 弱感情データから作成した英語の感情予測モデルの精度

感情	適合率	再現率	F 値
喜び	0.65	0.82	0.73
悲しみ	0.58	0.77	0.66
期待	0.53	0.81	0.64
驚き	0.45	0.74	0.56
怒り	0.39	0.53	0.45
恐れ	0.40	0.59	0.48
嫌悪	0.42	0.70	0.53
信頼	0.36	0.69	0.47
マクロ平均	0.47	0.71	0.56

### 3.3.2 中感情データから作成した感情予測モデルの精度

表 7 に日本語の感情予測モデルの感情ごとの精度を、表 8 に英語の精度を示す。表 7 のように、日本語のモデルの感情ごとの精度は、全ての感情に対して、適合率よりも再現率の方が高かった。感情ごとに精度を見ると、最も適合率が高い感情は喜びで **0.69**、最も再現率が高い感情は悲しみで **0.78**、最も F 値が高い感情は悲しみで **0.71** であった。また、最も適合率が低い感情は怒りで **0.35**、最も再現率が低い感情は恐れで **0.53**、最も F 値が低い感情は怒りで **0.43** であった。表 8 により、英語モデルの感情ごとの精度についても、日本語モデルと同様に、適合率よりも再現率の方が高かった。感情ごとに精度を見ると、最も適合率が高い感情は喜びで **0.65**、最も再現率が高い感情も喜びで **0.82**、最も F 値が

表 7 中感情データから作成した日本語の感情予測モデルの精度

感情	適合率	再現率	F 値
喜び	0.79	0.53	0.63
悲しみ	0.64	0.59	0.62
期待	0.61	0.55	0.58
驚き	0.52	0.51	0.52
怒り	0.60	0.22	0.32
恐れ	0.45	0.35	0.39
嫌悪	0.48	0.54	0.50
信頼	0.52	0.31	0.39
マクロ平均	0.58	0.45	0.49

表 8 中感情データから作成した英語の感情予測モデルの精度

感情	適合率	再現率	F 値
喜び	0.76	0.63	0.69
悲しみ	0.65	0.58	0.61
期待	0.59	0.64	0.61
驚き	0.52	0.51	0.51
怒り	0.51	0.29	0.37
恐れ	0.47	0.34	0.40
嫌悪	0.50	0.50	0.50
信頼	0.45	0.43	0.44
マクロ平均	0.56	0.49	0.52

高い感情は悲しみで **0.73** であった。また、最も適合率が低い感情は信頼で **0.36**、最も再現率が低い感情は怒りで **0.53**、最も F 値が低い感情は怒りで **0.45** であった。中感情データから作成した日本語のモデル全体のマクロ平均適合率は **0.49**、マクロ平均再現率は **0.67**、マクロ平均 F 値は **0.56** であり、英語のモデルのマクロ平均適合率は **0.47**、マクロ平均再現率は **0.71**、マクロ平均 F 値は **0.56** であった。

### 3.3.3 強感情データから作成した感情予測モデルの精度

表 9 に日本語の感情予測モデルの感情ごとの精度を、表 10 に英語の精度を示す。表 9 のように、日本語のモデルの感情ごとの精度は、全ての感情に対して、適合率よりも再現率の方が高かった。感情ごとに精度を見ると、最も適合率が高い感情は喜びで **0.69**、最も再現率が高い感情は悲しみで **0.78**、最も F 値が高い感情は悲しみで **0.71** であった。また、最も適合率が低い感情は怒りで **0.35**、最も再現率が低い感情は恐れで **0.53**、最も F 値が低い感情は怒りで **0.43** であった。表 10 により、英語モデルの感情ごとの精度についても、日本語モデルと同様に、適合率よりも再現率の方が高かった。感情ごとに精度を見ると、最も適合率が高い感情は喜びで **0.65**、最も再現率が高い感情も喜びで **0.82**、最も F 値が高い感情は悲しみで **0.73** であった。また、最も適合率が低い感情は信頼で **0.36**、最も再現率が低い感情は怒りで **0.53**、最も F 値が低い感情は怒りで **0.45** であった。強感情データから作成した日本語のモデル全体のマクロ平均適合率は **0.49**、マクロ平均再現率は **0.67**、マクロ平均 F 値は **0.56** であり、英語のモデルのマクロ平均適合率は **0.47**、マクロ平均再現率は **0.71**、マクロ平均 F 値は **0.56** であつ

表 9 日本語の感情予測モデルの感情ごとの精度

感情	適合率	再現率	F 値
喜び	0.86	0.36	0.50
悲しみ	0.77	0.25	0.38
期待	0.72	0.17	0.28
驚き	0.56	0.28	0.37
怒り	0.57	0.15	0.24
恐れ	0.55	0.10	0.17
嫌悪	0.64	0.16	0.26
信頼	0.50	0.23	0.32
マクロ平均	0.65	0.21	0.43

表 10 英語の感情予測モデルの感情ごとの精度

感情	適合率	再現率	F 値
喜び	0.83	0.40	0.54
悲しみ	0.67	0.50	0.57
期待	0.73	0.27	0.40
驚き	0.58	0.32	0.42
怒り	0.55	0.19	0.28
恐れ	0.47	0.21	0.29
嫌悪	0.61	0.28	0.38
信頼	0.54	0.21	0.30
マクロ平均	0.62	0.30	0.43

表 11 日本語の感情予測モデルの精度一覧

	弱感情データ	中感情データ	強感情データ
適合率	0.49	0.58	0.65
再現率	0.67	0.45	0.21
F 値	0.56	0.49	0.43

表 12 英語の感情予測モデルの精度一覧

	弱感情データ	中感情データ	強感情データ
適合率	0.47	0.56	0.62
再現率	0.71	0.49	0.30
F 値	0.56	0.52	0.43

た。

全ての感情予測モデルの中で、最も精度が高いモデルは、日本語は弱感情データを使用し、エポック数 2 で構築したモデル、英語も弱感情データを使用し、エポック数 2 で構築したモデルであった。モデルの精度はそれぞれ日本語が、マクロ平均適合率は 0.49、マクロ平均再現率は 0.67、マクロ平均 F 値は 0.56、英語がマクロ平均適合率は 0.49、マクロ平均再現率は 0.67、マクロ平均 F 値は 0.56 であった。これらのモデルを使用して第 4 章で絵文字感情ベクトルを作成する。

## 第4章 絵文字感情の分散表現

### 4.1 絵文字付きデータセット

テキストと絵文字の相互関係を考慮し、絵文字の定量化を行うために、私たちは絵文字を含むテキストを使用し、絵文字感情ベクトルを生成した。図 8 に絵文字付きデータセットの作成方法を示す。Unicode に登録されている絵文字 1146 種類のうち、X(旧 Twitter)にて 2022 年 8 月 18 日から 9 月 23 日にかけて各絵文字を含むポストを収集し、整形したデータセットを使用する。本論文では、1146 種類の絵文字のうち、日本語と英語のポスト(ツイート)が 1000 件以上取得できた絵文字 1064 種類を対象とする。1064 種類の各絵文字に対するテキストから、Bad Words リスト<sup>iii</sup>を使用してセンシティブな内容や暴力的な表現を取り除く。それらの表現を取り除いたテキストを 1000 件ずつランダムサンプリングし、テキストと絵文字のデータセットとする。絵文字 1056 種類ごとにテキスト数 1000 件、総数 1056000 件の絵文字付きテキストデータセットを日英で作成した。

### 4.2 感情予測に基づく絵文字感情分散表現

絵文字付きテキストデータセットを感情予測モデルに入力し、出力結果から 8 次元のベクトルを得る。ここで使用する感情予測モデルは、3.3 章で最も精度がよかったモデルを使用する。次に、8 次元のベクトルから、絵文字別に感情の要素の平均を計算する。図 8 で、具体的に絵文字の例を挙げて絵文字感情ベクトルの作成方法について説明する。「🌀」の絵文字ラベルがついたテキストの場合、図 8 のように「🌀」の絵文字が付いたテキスト 1000 件を感情予測モデル

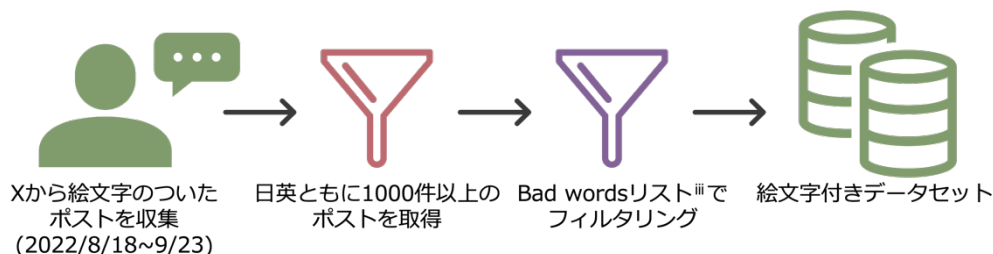


図 7 絵文字付きデータセットの作成方法

<sup>iii</sup> <https://github.com/LDNOOBW/List-of-Dirty-Naughty-Obscene-and-Otherwise-Bad-Words>

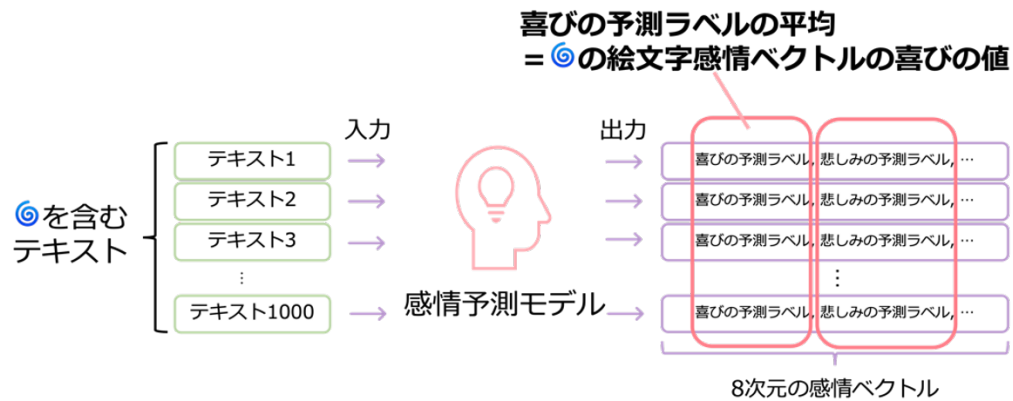


図 8 絵文字感情ベクトルの作成方法

に入力すると、1000 個の 8 次元ベクトルが得られる。得られた 1000 個のベクトルから感情の要素ごとに平均を計算する。得られた平均値を絵文字「㊦」の絵文字感情ベクトルの各感情の感情値と定義する。この手順を 1056 種類の絵文字に対して行い、絵文字感情ベクトルを生成する。

## 第5章 絵文字の文化差抽出

生成した絵文字感情ベクトルから日英間の絵文字の文化差を抽出する。文化差の抽出には、感情軸ごとの文化差抽出、感情対極に基づく文化差抽出、順序相関に基づく文化差抽出の3つの手法を段階的に用いる。各手法は、より広い視点から文化差を捉えることを目的としており、いずれかの段階で文化差が抽出された場合に、文化差のある絵文字とする。なお、各段階で抽出された文化差は、順に局所的文化差(第一種文化差)、対極的文化差(第二種文化差)、構造的文化差(第三種文化差)と定義する。いずれかに該当した絵文字の集合を、文化差のある絵文字集合と定義する。

### 5.1 感情軸に合わせたレーダーチャートによる視覚化

本論文で感情ラベルに用いた感情は Plutchik の基本 8 感情に基づいている。Plutchik の理論では、8 つの基本的な感情が存在し、人間の進化において生存に重要な役割を果たしてきたとされている。基本感情は、Plutchik の基本 8 感情の輪の対極に位置する感情に対立する性質を持っていると考えられている。喜びの感情と悲しみの感情は輪の対極に位置し、それらの感情は対になっている。本論文では、対になっている感情が互いに対立するという基本 8 感情の性質に着目して、絵文字感情の評価を行う。対になっている感情ペアはそれぞれ(喜び, 悲しみ), (恐れ, 怒り), (驚き, 期待), (信頼, 嫌悪)である。Plutchik の基本 8 感情の性質を用いて、絵文字 1056 件の絵文字感情ベクトルをレーダーチャート

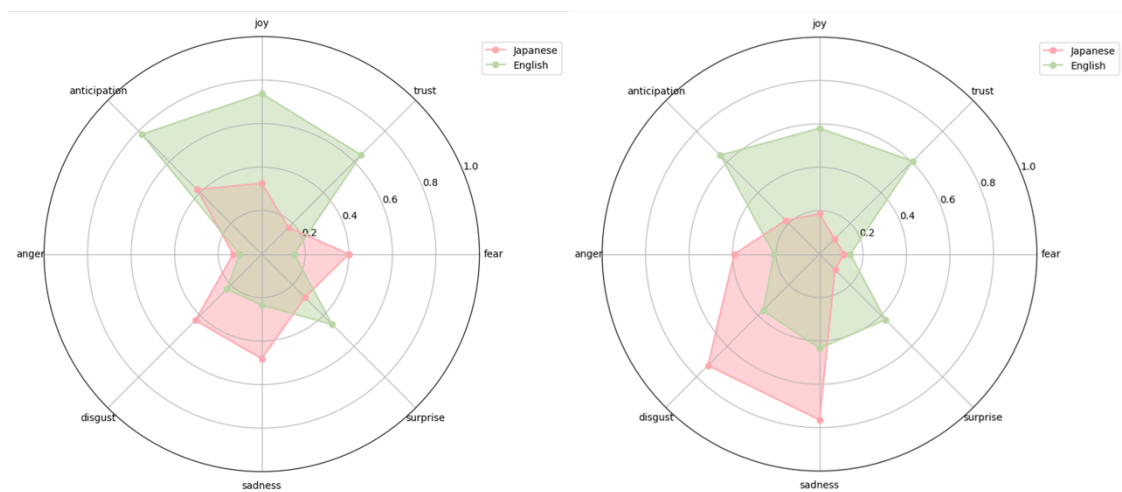


図 9 描画したレーダーチャートの例

にて描画する。図 9 に示すようなレーダーチャートで描画することで、絵文字感情ベクトルの感情の差を視覚的に捉えることが可能となる。

## 5.2 局所的文化差の抽出

本節では、各感情軸に着目し、絵文字の個別の感情の強さの違いに基づく文化差を抽出する。本手法では単一感情における局所的な差を捉えることを目的とし、抽出された文化差を局所的文化差(第一種文化差)と定義する。

具体的には、各絵文字について感情軸ごとに、日本語および英語の絵文字感情ベクトルにおける感情値の平均を比較し、t 検定に基づく仮説検定を行う。まず、8 次元の絵文字感情ベクトルを生成する過程で用いた平均をとる前の 8 次元の感情ベクトル集合から、感情軸ごとの平均を算出する。感情ベクトル数は絵文字 1 つに対し、日本語 1000 件、英語 1000 件ある。各感情値について、帰無仮説を「日本語群と英語群の平均値に差がない」として、t 検定を実施する。このとき、有意水準 5% で有意な文化差があるとする。しかし、検定に用いるサンプル数が多く、統計的有意性だけでは実質的な差の大きさを評価できないため、Cohen の効果量[11]を算出し、表 13 の基準である効果量の解釈が Medium, Large の絵文字を局所的文化差がある絵文字とする。

1056 件の絵文字を対象とした検定の結果、合計 293 件の絵文字に第一種文化差が抽出された。感情軸ごとの内訳は、喜び軸で 45 件、悲しみ軸で 59 個、期待軸で 115 件、驚き軸で 136 件、怒り軸で 50 件、恐れ軸で 32 件、嫌悪軸で 43 件、信頼軸では 110 件の表 14 に示す絵文字であった。

## 5.3 対極的文化差の抽出

5.2 章での文化差抽出手法では、感情間の方向性の違いを十分に捉えられない可能性がある。Plutchik による基本 8 感情の理論では、感情は円環上に配置され、対極に位置する感情同士は相対する性質を持つとされている。対極感情間の

表 13 Cohen の効果量の解釈

効果量 d	解釈
$d < 0.2$	Very small
$0.2 \leq d < 0.5$	Small
$0.5 \leq d < 0.8$	Medium
$d \geq 0.8$	Large











は、各感情ペアの前者の感情値から後者の感情値を引いて差を算出する。対極関係にある感情ペアは(喜び, 悲しみ), (恐れ, 怒り), (驚き, 期待), (信頼, 嫌悪)である。感情ベクトル数は絵文字1つに対し、日本語1000件、英語1000件ある。対になっている感情の値について、帰無仮説を「日本語群と英語群の平均値に差がない」として、t検定を実施する。このとき、有意水準5%で有意な文化差があるとする。しかし、検定に用いるサンプル数が多く、統計的有意性だけでは実質的な差の大きさを評価できないため、Cohenの効果を算出し、効果量の解釈がMedium, Largeの絵文字を対極的文化差(第二種文化差)がある絵文字とする。

表 15 第二種文化差のある絵文字

感情軸	第二種文化差のある絵文字	
(喜び, 悲しみ)	第二種のみ	
	第一種と重複	
(期待, 驚き)	第二種のみ	
	第一種と重複	
(怒り, 恐れ)	第二種のみ	
	第一種と重複	
(嫌悪, 信頼)	第二種のみ	
	第一種と重複	

表 15 に第二種文化差が抽出された絵文字一覧を示す。(喜び, 悲しみ)軸で有意な文化差が得られた絵文字は 43 件, (驚き, 期待)軸で有意な文化差が得られた絵文字は 28 件, (恐れ, 怒り)軸で有意な文化差が得られた絵文字は 27 件, (信頼, 嫌悪)軸で有意な文化差が得られた絵文字は 88 件であった。これらのうち, 多くの絵文字は第一種文化差と重複していたが, 単一感情軸では文化差が認められず, 第二種文化差として新たに抽出された絵文字は表 15 の 6 件(, , , , , )であった。

## 5.4 構造的文化差の抽出

5.2 章, 5.3 章では, 単一の感情軸と対極感情軸に基づき文化差を抽出したが, これらの文化差抽出手法では, 絵文字が示す感情全体の構造的な違いを十分に捉えられていない可能性がある。そこで本節では, 絵文字感情ベクトル全体に着目し, 感情構造に基づく文化差を抽出する。この手法によって抽出された文化差を構造的文化差(第三種文化差)と定義する。

具体的には, 二変数の順位データから相関を測定する指標であるスピアマンの順位相関係数を算出する。感情値の絶対的な大きさではなく, 相対的な順位関係を評価できる。絵文字感情ベクトルの各要素は順位データではないため, 各絵文字について 8 次元の感情値を大きい順に順位を付与する。これにより得られた順位を用いて日英の順位を比較し, スピアマンの順位相関係数を算出する。順位相関係数が低いほど, 絵文字全体の感情の構造が日英間で一致していないことを示す。






生成した絵文字感情ベクトルに対して順位相関係数を算出した結果, 相関係数が負の値を示す絵文字は 5 件 (, , , , ) 存在したが, 有意水準 5%( $p \leq 0.05$ )において統計的に有意な差は認められなかった。また, 順位相関係数が 0.5 以下の絵文字は 37 件であり, これらについても同様に統計的な有意差

表 16 第三種文化差のある絵文字








































第三種のみ	 ,  ,  ,  ,  ,  , 
第一種と重複	 ,  ,  ,  , 
第二種と重複	
第一種, 第二種と重複	 ,  ,  ,  ,  ,  ,  ,  ,  ,  ,  ,  ,  ,  ,  ,  ,  ,  ,  ,  ,  ,  ,  ,  ,  ,  , 

表 17 第三種文化差のみの絵文字

絵文字	相関係数	p 値
👉	0.405	0.320
🔪	0.190	0.651
😂	0.429	0.289
😬	0.452	0.260
🐱	0.347	0.399
😮	0.190	0.651
🤪	0.167	0.693

は確認されなかった。絵文字感情の構造的違いが比較的大きい絵文字候補を探索的に抽出するために、ヒューリスティックな基準として相関係数が 0.5 以下の絵文字を第三種文化差として設定した。表 16 に、相関係数が 0.5 以下の 37 件の第三種文化差のある絵文字と、第一種、第二種文化差との重複関係を示す。相関係数が 0.5 以下の 37 件の絵文字のうち、第一種、第二種文化差と重複する絵文字は 30 件あったが、第三種文化差として新たに文化差が抽出された絵文字は、表 17 に示す 7 件(👉, 🔪, 😂, 😬, 🐱, 😮, 🤪)であった。

最後に、提案手法によって文化差が抽出された絵文字数の関係を整理するため、第一種～第三種の抽出結果の重なりをベン図として図 11 に示す。図 11 よ

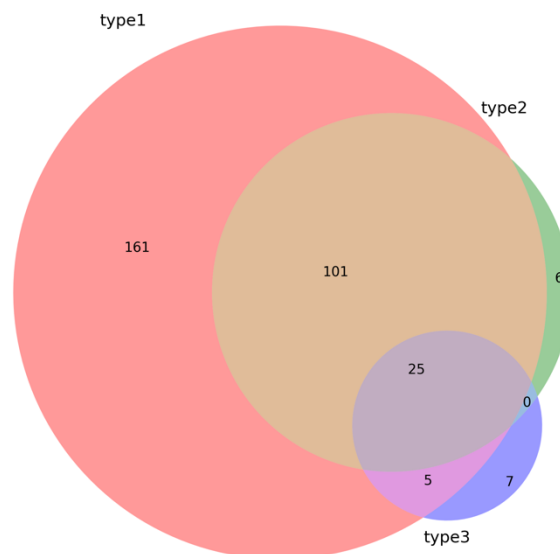


図 11 提案手法における文化差抽出結果の文化差種類別重なり

り，第一種文化差として抽出された絵文字数が最も多く，提案手法における文化差の大部分が第一種文化差であることがわかる．また，第二種，第三種文化差では第一種文化差と共通して抽出される絵文字が多い一方で，第二種，第三種文化差のみで抽出された絵文字も存在しており，第一種文化差では捉えきれなかった文化差を補完的に抽出していることが確認できる．

## 第6章 評価

主観的な評価には、人手でのアンケートを実施した。日本語話者と英語話者の感情表現の異なる絵文字を統計的に文化差が得られた絵文字と比較する。具体的には、抽出した文化差のある絵文字が、人が感じる文化差とどの程度整合するのかどうかを重複にて評価する。人手ベースの文化差は、人手での絵文字感情アノテーションにより作成した手作業の絵文字感情ベクトルを用いて算出した。

### 6.1 人手での絵文字感情アノテーション

人手でのアンケートは、20代から30代の日本語話者10名と、英語話者6名に絵文字のアノテーション作業を行ってもらった。絵文字のアノテーション作業の具体的な内容は、絵文字についてどんな感情を表現するときを使用するかどうかを、Plutchikの基本8感情の輪と同じ感情の強度(0, 1, 2, 3)から選択してもらった。さらに自信のない回答にはチェックをつけてもらった。図12に実際の絵文字アノテーション作業に対する説明を、図13に実際のアノテータによる人手での感情ラベリングの例を示す。アノテータは、図12の作業説明を読み、

各絵文字について、どんな感情を表現するときを使用するか当てはまるものを全て教えてください。またその感情がどれくらいの強さなのか、以下の各感情の説明を参考に教えてください。回答は、「0」、「1」、「2」、「3」いずれかの数字を表に入れてください。

自分の回答に対して、自信のない感情には「○」をつけてください。

《回答例》

	喜び	喜び_自信度	悲しみ	悲しみ_自信度	恐れ	恐れ_自信度	怒り	怒り_自信度	驚き	驚き_自信度	期待	期待_自信度	嫌悪	嫌悪_自信度	信頼	信頼_自信度
☺	1	○	2		0		1	○	0		3		0		1	

《各感情の説明》

本研究で取り扱っている感情はPlutchikの基本8感情に基づいたもので、感情の種類や強さについて定義されています。感情の強さは、左から順番に0(無)、1(弱)、2(中)、3(強)です。例えば「喜び」の感情では、「無し」が0(無)「平穏」が1(弱)、「喜び」が2(中)、「歓喜」が3(強)を示します。

- ・無(0) < 平穏(1) < 喜び(2) < 歓喜(3)
- ・無(0) < 哀愁(1) < 悲しみ(2) < 悲嘆(3)
- ・無(0) < 不安(1) < 恐れ(2) < 恐怖(3)
- ・無(0) < イライラ(1) < 怒り(2) < 激怒(3)
- ・無(0) < 放心(1) < 驚き(2) < 驚嘆(3)
- ・無(0) < 関心(1) < 期待(2) < 警戒(3)
- ・無(0) < うんざり(1) < 嫌悪(2) < 強い嫌悪(3)
- ・無(0) < 受容(1) < 信頼(2) < 敬愛(3)

図 12 実施した人手での絵文字感情アンケートの設問

	喜び	喜び_自信度	悲しみ	悲しみ_自信度	恐れ	恐れ_自信度	怒り	怒り_自信度	驚き	驚き_自信度	期待	期待_自信度	嫌悪	嫌悪_自信度	信頼	信頼_自信度
🌀	0		1	○	0		0		0		0		0		0	
🎨	1		0		0		0		0		0		0		1	○
🖌️	0		3		0		0		0		0		1	○	0	
🖼️	2		0		0		0		0		2		0		0	
🖱️	2		0		0		0		0		2		0		0	
🖨️	2		0		0		0		0		2		0		0	
🖨️	2		0		0		0		0		2		0		0	
🖨️	2		0		0		0		0		2		0		0	
🖨️	2		0		0		0		2		3		0		2	
🖨️	1		0		0		0		1	○	0		0		1	○

図 13 実際のアノテータによる回答

それぞれの絵文字を見て、各感情に対する 4 段階の強度の解答と自信がない感情に○を記入している。具体的に図○では、「🌀」の絵文字について、アノテータは喜び、恐れ、怒り、驚き、期待、嫌悪、信頼の感情に「0」、悲しみの感情に「1」とアノテーションしたが、悲しみの感情に対する自信度に「○」がついているため、「🌀」の悲しみの感情に対するアノテーションへの自信がないことを示している。今回アノテーション作業の対象とした絵文字数は 1056 件である。

また、モデルを使用して作成した絵文字感情ベクトルと絵文字アノテーショ

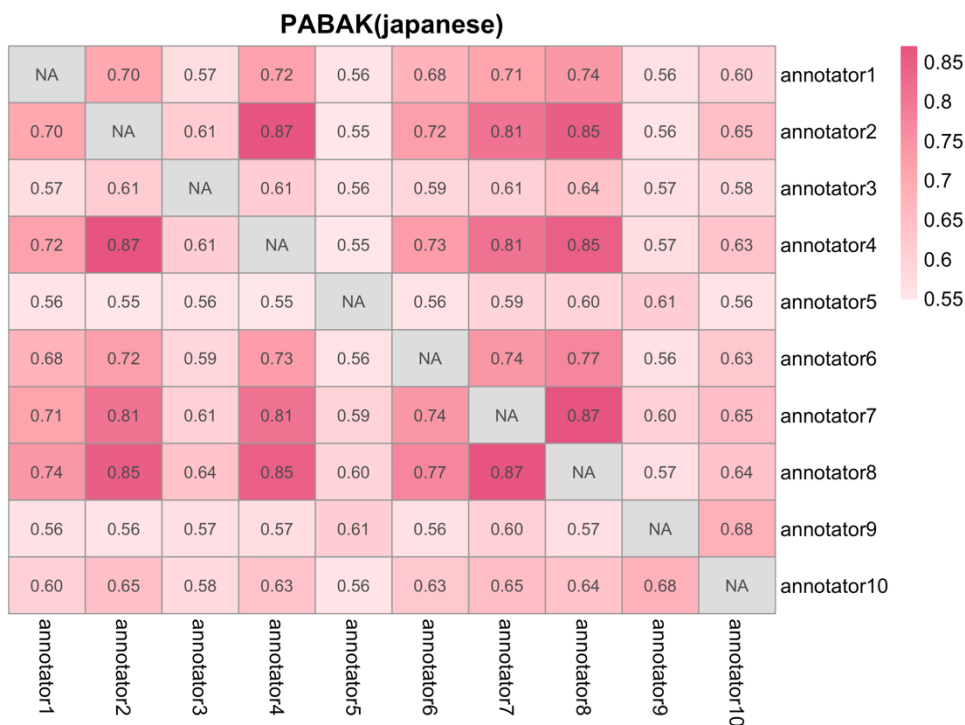


図 14 日本語でのアノテータ間の PABAK

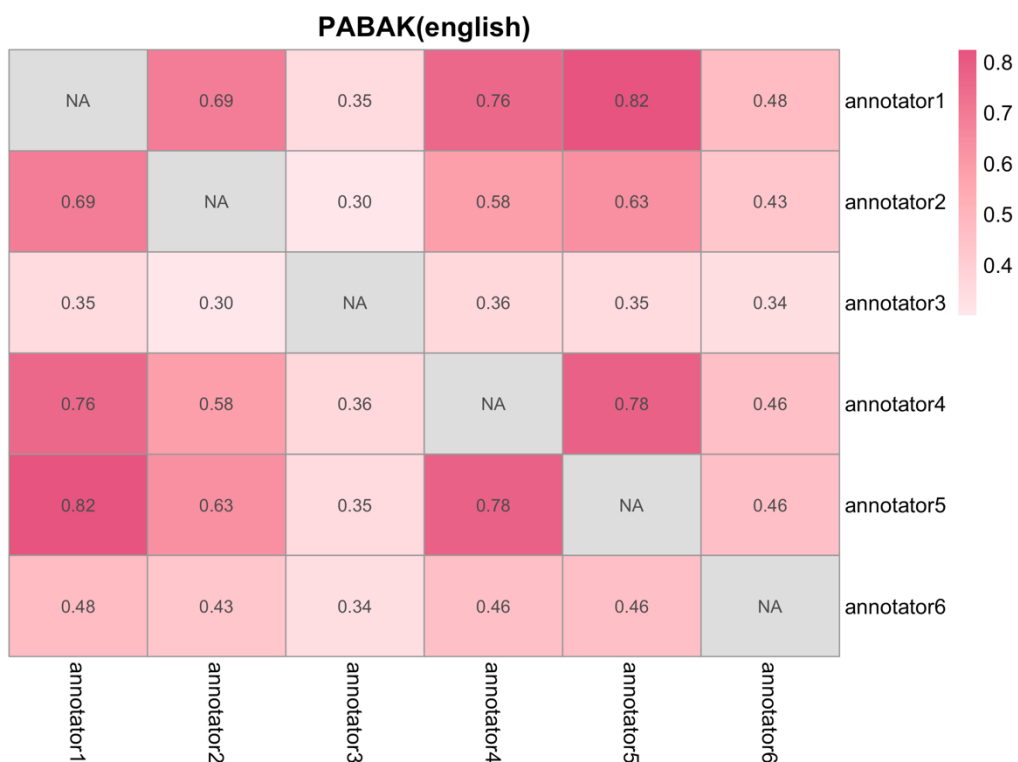


図 15 英語でのアノテータ間の PABAK

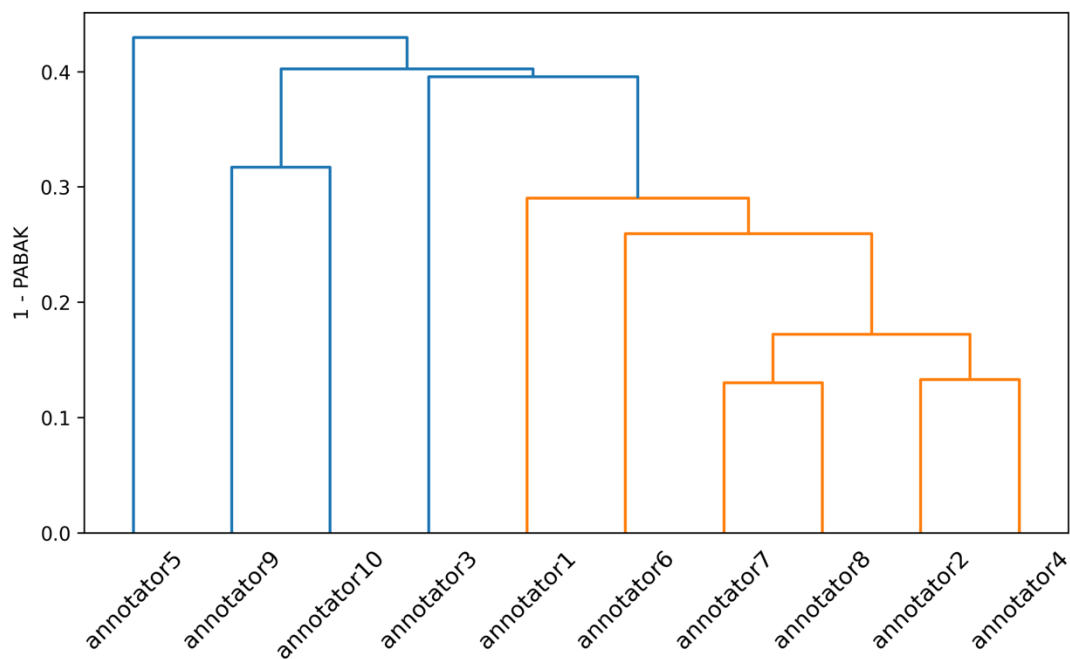


図 16 日本語アノテータ群の PABAK での階層型クラスタリング結果  
業の対象とした絵文字数は 1056 件である。

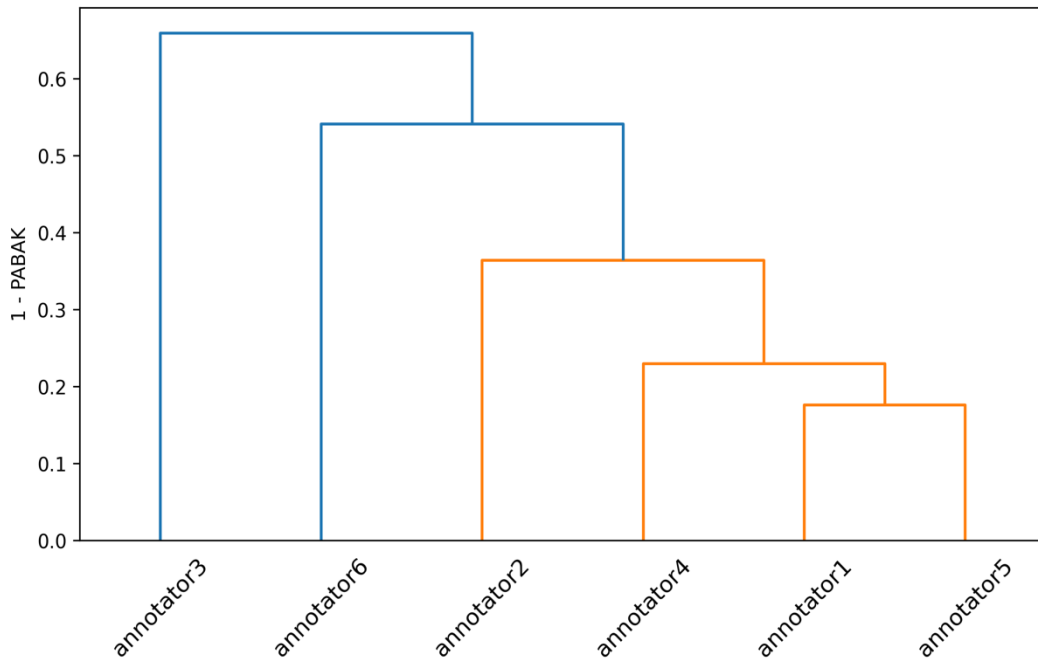


図 15 英語アノテータ群の PABAK での階層型クラストリング結果

また、モデルを使用して作成した絵文字感情ベクトルと絵文字アノテーションデータを比較するために、感情強度をモデルベースと揃える必要がある。そこで、感情強度が 0 のものは 0, 1 以上のものはすべて 1 に置換したデータを評価に使用した。さらに、日英のアノテーション作業者のデータを、絵文字ごとに集計し、それぞれの感情について平均をとった。これを人手ベースの絵文字感情ベクトルと称し、モデルベースの絵文字感情ベクトルと比較することで、モデルベースの絵文字感情ベクトルの妥当性と絵文字の文化差の評価を行う。

しかしながら、人手での絵文字への評価には揺らぎが生じる。そのため、人手でのアノテーション結果の信頼性を、PABAK[11]を算出することで検証した。一般的に、人手での評価の信頼性の評価には Kappa 係数が用いられるが、感情強度が「0」で一致する場合は非常に多い本アノテーションの評価では、Kappa 係数が評価の一致度を過小に評価される可能性があった。そこで、クラスの出現頻度による影響を補正できる PABAK を使用し評価を行った。

今回のアノテータは日本語話者 10 名、英語話者 6 名である。各アノテータ間での PABAK を算出した結果を日英それぞれ図 14, 図 15 のヒートマップにて示す。図 14 から、日本語アノテータ間の PABAK は全てのペアに対して 0.5 より

も大きく、アノテータ間の回答が比較的一致していることが確認でき、アノテータ間同士の回答の一致度が非常に高いペアも複数見られる。図 15 の英語アノテータ間の PABAK を見ると、日本語アノテータ群と比べると PABAK が小さく、アノテータによっては別のアノテータと回答にばらつきがあることがわかる。そこで、アノテータ 2 者間の PABAK を距離尺度として、階層型クラスタリングを行い、アノテータ間の評価の一致を確認した。階層型クラスタリングの結果について、日本語アノテータ群を図 16 に、英語アノテータ群を図 15 に示す。日本語アノテータ群では、図中の右側のアノテータ 6 名が同一クラスタを形成しており、図 14 のヒートマップを見ても高い PABAK である。また、英語アノテータ群では、図 15 中の右側の 4 名が同一クラスタを形成しており、回答の一致度が高いことがわかる。これらの 4 名についても、図 15 のヒートマップでアノテータ間の PABAK から PABAK が高いことが確認された。日本語アノテータ群では、PABAK が 0.5 を超えるペアが他にも存在していたが、個々のペアにおける評価の一致度の高さよりも、集団として一貫した評価が行われていることを重視し、同一クラスタに所属するアノテータ軍を信頼性の高いアノテータ群として採用した。本研究では、日本語アノテータ 6 名、英語アノテータ 4 名分

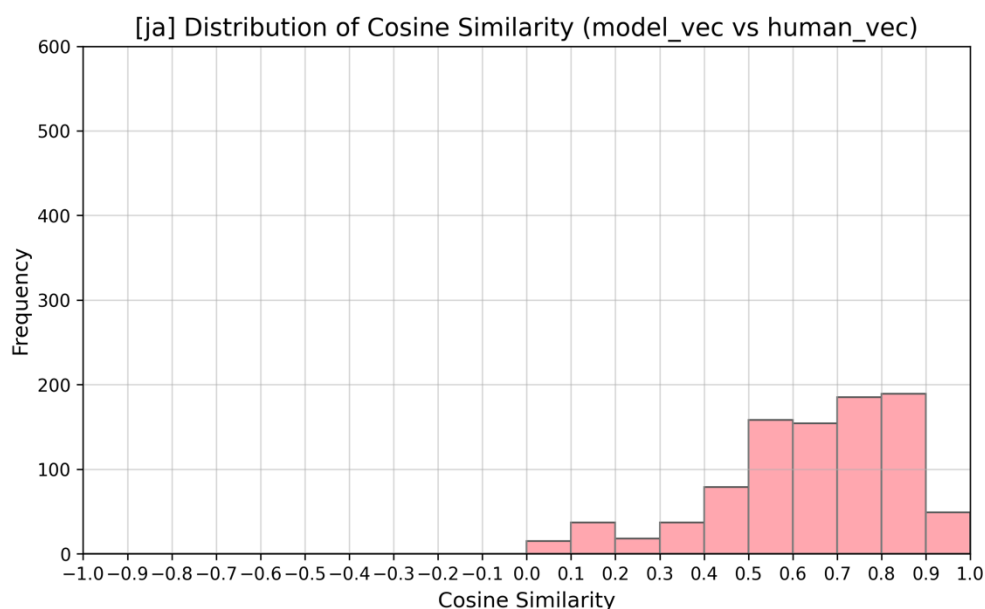


図 16 日本語のモデルベースの絵文字感情ベクトルと人手ベースの絵文字感情ベクトルのコサイン類似度の分布

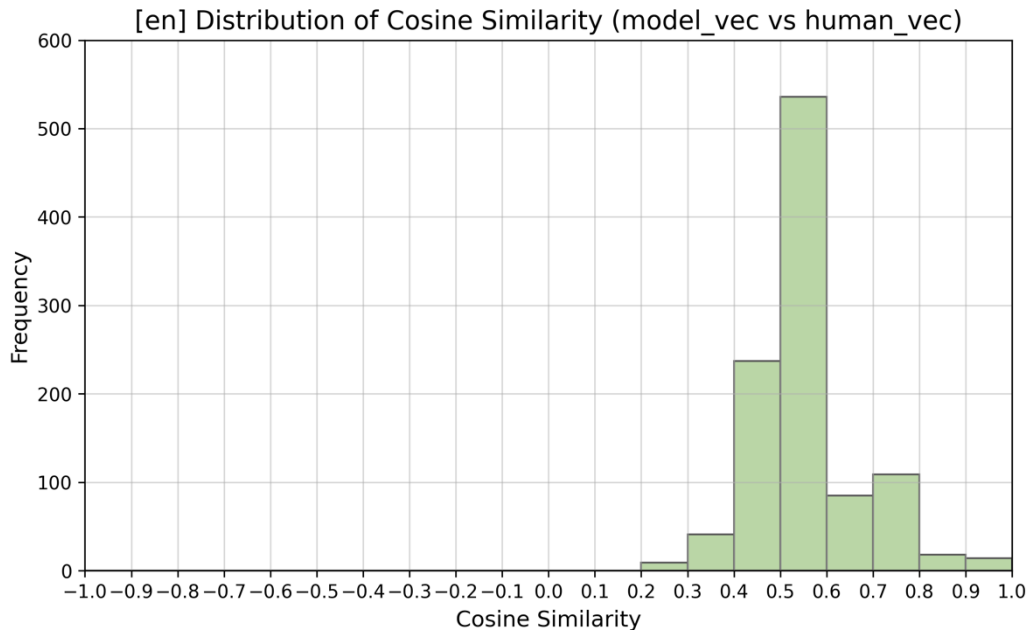


図 17 英語のモデルベースの絵文字感情ベクトルと  
人手ベースの絵文字感情ベクトルのコサイン類似度の分布

の評価結果を正解データとして今回は用いることとする．この日本語アノテータ 6 名，英語アノテータ 4 名分の PABAK の平均はそれぞれ 0.71，0.77 であった．

## 6.2 絵文字感情分散表現の正当性

まず，第 4 章で作成した絵文字感情分散表現について評価を行う．評価には，コサイン類似度，を算出することで，人が感じる絵文字への感情が提案手法で作成した絵文字感情の分散表現と数値的に，構造的に一致しているかを評価する．どの手法も，第 4 章で作成した絵文字感情の分散表現と人手でのアノテーションデータから作成した絵文字感情の分散表現から算出する．今後，各分散表現のことをモデルベースの絵文字感情ベクトル，人手ベースの絵文字感情ベクトルと呼ぶ．

モデルベースと人手ベースの絵文字感情ベクトルから算出したコサイン類似度の分布を表したヒストグラムを日英それぞれ図 16，図 17 に示す．日本語のコサイン類似度は図 16 を見ると，コサイン類似度が負の絵文字感情ベクトルはなく，全ての絵文字についてモデルと人手のベクトルが比較的同じ角度を向いていることがわかる．また，半数以上の絵文字のコサイン類似度が 0.5 以上であ

り、8.8以上の絵文字も200個以上存在している。図19を見ると、日本語と同様に英語のベクトルのコサイン類似度が負の絵文字はなく、0.5以上の絵文字が多い。最も標本数が多いコサイン類似度は0.5~0.6であり、日本語のベクトルの方が英語に比べ、ベクトル同士の角度の観点では人が実際に抱く感情を再現できていると言える。

### 6.3 人手による文化差のある絵文字の定義







本節では、人が感じる文化差のある絵文字を人手ベースの絵文字感情ベクトルを用いて定義する。人手による文化差の判定には、抽出観点の違いに基づいた3手法を用い、それぞれ人手の第一種文化差、第二種文化差、第三種文化差と呼ぶ。いずれかに該当した絵文字の集合を、人手による文化差のある絵文字集合と定義する。

まず、第一種文化差では、各感情軸に着目し、絵文字の個別の感情の強さの違いに基づく文化差を抽出する。具体的には、各絵文字について感情軸ごとに、日本語および英語の人手アノテーション値に対して Mann-Whitney の U 検定を行い、日英の二群間で分布に差があるかを検定する。帰無仮説は「日本語群と英語群の分布に差がない」とし、有意水準  $5\%(p \leq 0.05)$  で判定する。さらに、統計的有意性のみでは差の大きさを評価できないため、効果量として Cliff's del $\tau$ [13] を算出し、その解釈が effsize パッケージで示されている閾値が Medium または Large[14] に該当する場合に、第一種文化差において実質的な差があるとみなす。合計32件の絵文字が人手の第一種文化差ありとして抽出された。感情軸ごとの内訳は、喜び軸で16件、悲しみ軸で3件、期待軸で3件、驚き軸で4件、怒り軸で1件、恐れ軸で0件、嫌悪軸で5件、信頼軸では1件の表18に示す絵文字であった。

第一種文化差に加え、第二種文化差では、対極関係にある感情軸の差分に着目し、感情の方向性やバランスの違いに基づく文化差を抽出する。具体的には、対極関係にある感情ペアは、(喜び, 悲しみ), (恐れ, 怒り), (驚き, 期待), (信頼, 嫌悪)の4組とする。各絵文字について、感情ペアごとに前者の感情値から後者の感情値を引くことで差を算出する。得られた差に対して Mann-Whitney の U 検定を行い、日英の二群間で分布に差があるかを検定する。帰無仮説は「日本語群と英語群の分布に差がない」とし、有意水準  $5\%(p \leq 0.05)$  で判定する。統計的有意性のみでは差の大きさを評価できないため、効果量として Cliff's del $\tau$

表 18 人手で第一種文化差のある絵文字

感情軸	人手で第一種文化差のある絵文字
喜び	
悲しみ	
期待	
驚き	
怒り	
恐れ	
嫌悪	
信頼	

を算出し、その解釈が **Medium** または **Large** に該当する場合に、人手の第二種文化差において実質的な差があるとみなす。(喜び, 悲しみ)軸で有意な文化差が得られた絵文字は 19 件, (驚き, 期待)軸で有意な文化差が得られた絵文字は 1 件, (恐れ, 怒り)軸で有意な文化差が得られた絵文字は 2 件, (信頼, 嫌悪)軸で有意な文化差が得られた絵文字は 5 件であった。これらのうち、複数の絵文字は第一種文化差と重複していたが、単一感情軸では文化差が認められず、第二種文化差として新たに抽出された絵文字は表 19 の 6 件(, , , , , )であった。

さらに、人手の第一種文化差、第二種文化差に加え、第三種文化差では、絵文字感情ベクトル全体に着目し、感情構造に基づく文化差を抽出する。具体的には、

表 19 人手で第二種文化差のある絵文字

感情軸	第二種文化差のある絵文字	
喜び—悲しみ	第二種のみ	
	第一種と重複	
期待—驚き	第二種のみ	
	第一種と重複	
怒り—恐れ	第二種のみ	
	第一種と重複	
嫌悪—信頼	第二種のみ	
	第一種と重複	

表 20 人手で第三種文化差のある絵文字

<p>第三種のみ</p>	
<p>第一種と重複</p>	
<p>第二種と重複</p>	
<p>第一種, 第二種 と重複</p>	

各絵文字について 8 次元の感情値を大きい順に並べ、順位を付与することで感情順位構造を得る。各絵文字について 8 次元の感情値を大きい順に順位を付与する。これにより得られた順位を用いて日英の順位を比較し、スピアマンの順位相関係数を算出する。順位相関係数が 0.5 以下の場合、人手の第三種文化差において実質的な文化差があるとみなす。相関係数が負の値を示す絵文字は 146 件存在したが、有意水準 5%( $p \leq 0.05$ )において統計的に有意な差は認められなかった。また、順位相関係数が 0.5 以下の絵文字は 416 件であり、これらについても同様に統計的な有意差は確認されなかった。相関係数が 0.5 以下の 416 件の絵文字のうち、第一種、第二種文化差と重複する絵文字は 23 件あったが、第三種文化差として新たに抽出された絵文字は表 20 に示す 393 件であった。

最後に、人手ベースで文化差が抽出された絵文字数の関係を整理するため、第一種~三種の抽出結果の重なりをベン図として図 18 に示す。図 18 より、人手で第三種文化差として抽出された絵文字数が最も多く、文化差の大部分が感情の順位構造の違いに基づいて捉えられていることがわかる。また、第一種および第二種で抽出された絵文字数は比較的少ないものの、第三種と共通して抽出される絵文字も存在しており、複数の観点から文化差が示唆される絵文字が含ま

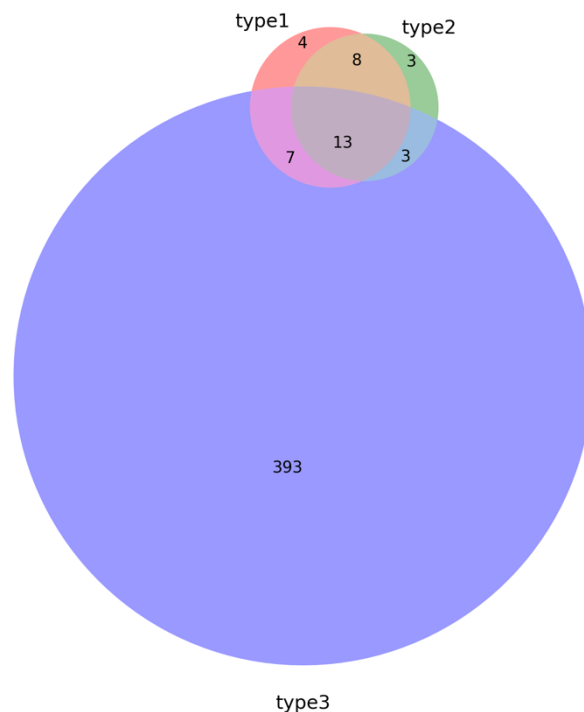


図 18 人手ベースでの文化差抽出結果の文化差種類別重なり

まれていることが確認できる。さらに、第一種または第二種のみで抽出された絵文字も少数存在し、単一感情軸や対極感情軸に基づく文化差もそれぞれ補完的に抽出されていることがわかる。

## 6.4 提案手法によって抽出された文化差の妥当性

提案手法によって抽出された文化差絵文字集合の妥当性を検証する。具体的には、6.3 で定義した人手ベースの文化差のある絵文字集合を基準とし、提案手法による抽出結果との一致度を評価する。評価にあたっては、提案手法で抽出した文化差のある絵文字集合と、人手ベースの文化差のある絵文字集合の比較を行い、真陽性(TP)、偽陽性(FP)、偽陰性(FN)を以下のように定義する。

- TP：提案手法と人手の両方で文化差ありとして抽出された絵文字
- FP：提案手法では抽出されたが、人手では文化差ありと判定されなかった絵文字
- FN：人手では文化差ありと判定されたが、提案手法では抽出されなかった絵文字

これらに基づき、提案手法の文化差抽出の妥当性を適合率、再現率、F 値によって定量的に評価する。適合率は提案手法が抽出した絵文字のうち人手と一致した割合を示し、再現率は人手で抽出された文化差のある絵文字を提案手法がどの程度網羅できたかを示す。

表 21 に、文化差種類別、また全体の文化差抽出精度を示す。提案手法が抽出した文化差のある絵文字集合が 305 件、人手ベースの文化差のある絵文字集合が 431 件であり、そのうち TP が 137 件、FP が 168 件、FN が 294 件存在した。それぞれに含まれる絵文字の全件の一覧は付録 A.1 に示す。この結果から妥当性を

表 21 文化差種類別・全体の文化差抽出精度

	第一種	第二種	第三種	全体
提案手法で文化差あり	292	132	37	305
人手ベースで文化差あり	32	27	416	431
TP	8	4	21	137
FP	284	128	16	168
FN	24	23	395	294
適合率	0.027	0.030	0.57	0.45
再現率	0.25	0.15	0.051	0.32
F 値	0.049	0.050	0.093	0.37

算出した結果、適合率は **0.45**、再現率は **0.32**、F 値は **0.37** であった。また、文化差種類別では、第一種文化差で提案手法が抽出した文化差のある絵文字集合が **292** 件、人手ベースの文化差のある絵文字集合が **32** 件であり、そのうち TP が **8** 件、FP が **284** 件、FN が **24** 件存在し、適合率、再現率、F 値はそれぞれ **0.027**、**0.25**、**0.049** であった。第二種文化差で提案手法が抽出した文化差のある絵文字集合が **132** 件、人手ベースの文化差のある絵文字集合が **27** 件であり、そのうち TP が **4** 件、FP が **128** 件、FN が **23** 件存在し、適合率、再現率、F 値はそれぞれ **0.030**、**0.15**、**0.050** であった。第三種文化差で提案手法が抽出した文化差のある絵文字集合が **37** 件、人手ベースの文化差のある絵文字集合が **416** 件であり、そのうち TP が **21** 件、FP が **16** 件、FN が **395** 件存在し、適合率、再現率、F 値はそれぞれ **0.57**、**0.051**、**0.093** であった。

## 第7章 考察

### 7.1 感情予測モデルの特徴と限界

第3章で述べた，感情予測モデルの構築手法を踏まえ，その設計が実際の予測性能や分析結果に与えた影響について考察する．まず，構築した感情予測モデルの特徴について述べる．1点目の特徴は，テキストに対して複数の感情が付与されうる状況を想定し，感情予測をマルチラベル分類問題として定式化した点である．本研究では，8感情それぞれに対して感情の有無の二値判定を同時に学習する構成を採用している．この構成により，単一の感情だけでなく，複数の感情の共起や感情の有無を分析に反映することが可能となる．2点目の特徴は，モデルの学習時，予測において複数のパラメータを調整した点である．具体的には，言語ごとにそれぞれ最適なエポック数を設定し，学習が未学習，過学習にならないよう調整を行なった．さらに，マルチラベル分類において，感情ごとに出現頻度や予測値の分布が異なるため，各感情に対して一律の閾値を適用するのではなく，検証データにおいてF値が最大となるような最適閾値を感情ごとに設けている．この設計により，感情間の不均衡や予測値の分布の違いを考慮した柔軟な予測が可能となる．

一方で，本研究で用いた感情予測モデルにはいくつかの制約が存在する．まず，トレーニングデータにおける感情ごとの正例数を表22に示す．表22のように，「喜び」，「悲しみ」，「期待」の感情では正例数が多いが，「怒り」，「恐れ」，「嫌悪」，「信頼」では正例数が少ない傾向が見られた．特に怒りは他の感情と比

表 22 学習データにおける感情ごとの正例数

感情	正例数
喜び	13956
悲しみ	12193
期待	12999
驚き	9997
怒り	4445
恐れ	6086
嫌悪	7156
信頼	7526

表 23 テストデータにおける感情別の正例数と予測正例数

感情	テストデータ	日本語モデル	英語モデル
喜び	1741	1867	2185
悲しみ	1514	2016	2029
期待	1606	2269	2453
驚き	1221	1628	2004
怒り	576	929	774
恐れ	761	1020	1123
嫌悪	895	1705	1481
信頼	914	1098	1760

較して著しく正例数が少なく、感情間で正例数に不均衡が見られる。このような学習データの感情ごとの正例数の分布は、感情ごとの予測性能に影響を与える可能性がある。実際に、日英の感情予測モデルの評価結果を確認すると、表 5、表 6 のように正例数の多い「喜び」、「悲しみ」、「期待」の感情は日英ともに比較的高い F 値が得られている。その一方で、正例数の少ない感情では F 値が低い傾向が見られた、特に正例数が少ない「怒り」では、日英ともに全感情の中で最も低い F 値であった。また、感情別のテストデータにおける正例数とモデルが予測した正例数を表 23 に示す。表 23 を確認すると、全ての感情で正解ラベル数よりも多くの正例の予測が行われていた。このことは、本研究の感情予測モデルが感情を比較的広く拾い上げることを示しており、表 5、表 6 の再現率が適合率を上回る結果と整合している。一方で、このようなモデルの傾向は特に正例数の少ない感情において正例の誤抽出を増加させ、適合率の低下の要因になっていると考えられる。

## 7.2 絵文字感情ベクトルの特徴

本研究で用いた二種類の絵文字感情ベクトルについて、それぞれの構築方法や特徴から考察する。感情予測に基づく絵文字感情ベクトルを提案手法、人手アノテーションによる絵文字感情ベクトルを比較基準と仮定している。

### 7.2.1 感情予測に基づくモデルベースの絵文字感情ベクトル

提案手法である感情予測に基づく絵文字感情ベクトルは、感情予測モデルを用いて、実際の X の投稿で絵文字が含まれるテキストから作成したものである。具体的には、特定の絵文字を含むポストを収集し、各ポストに対して感情予測を

行なった上で、得られた感情値を平均することで、8次元の絵文字感情ベクトルを表現している。

この手法では、絵文字単体ではなく、絵文字が使用された文脈を入力として絵文字に対する感情を推定している点が特徴である。そのため、この絵文字感情ベクトルでは、絵文字の実際の使用文脈に込められた感情を反映した表現ができると考えられる。また、大量の使用データを基に構築している点も大きな特徴である。1つの絵文字につき1000件のポストを用いて感情値を集約している。そのため、個々のポストに含まれる個人のノイズは平均化することによって緩和され、客観的な定量化された表現ができていると考えられる。

一方で、感情予測に基づく絵文字感情ベクトルは、感情予測モデルの性能やポストを収集した時期に依存する。この点については、文化差が抽出された絵文字の評価・分析についての考察で留意する必要がある。

### 7.2.2 人手アノテーションによる人手ベースの絵文字感情ベクトル

人手アノテーションによる絵文字感情ベクトルは、提案手法での絵文字が抱く感情や文化差抽出結果の現実世界との妥当性を確認するために、絵文字に対する主観的な解釈に基づいて構築されたものである。各絵文字に対して、複数のアノテータが8感情について評価を行い、その結果を集約することで8次元の絵文字感情ベクトルを表現している。

人手アノテーションでは、絵文字単体を提示し、「絵文字をどのような感情の時に使用するか」をアノテータが想起し評価を行う形式を採用している。そのため、人手アノテーションによる絵文字感情ベクトルは、実際の使用文脈というよりも、絵文字に対する感情イメージを反映しやすいという特徴があると考えられる。本研究で分析に使用した人手アノテーション結果は、アノテータ間で一定の基準の一致が確認されたものを使用している。さらに、分析に使用したデータは、日本語話者6名、英語話者4名分であり、提案手法と比較するとデータ量が非常に限定的である。一定の基準を設け、アノテーションの一致を確認したとはいえ、完全にアノテータ間で評価にばらつきがないとは言い切れず、絵文字に対する評価は個人の解釈に依存するため、完全な正解として使用することは難しいと考えられる。

### 7.2.3 両手法による絵文字感情ベクトルの違い

提案手法の感情予測に基づく絵文字感情ベクトルと人手アノテーションによる絵文字感情ベクトルの間には、構築手法によって表現している感情の違いが

存在する。前者の感情予測に基づく絵文字感情ベクトルは、実際の  $X$  でのポストをもとに文脈に依存した感情傾向を表現するのに対し、後者の人手アノテーションによる絵文字感情ベクトルは、絵文字単体に対する感情傾向を表現している。

また、両者には使用しているデータ量やデータ収集時期にも違いがある。前者は、1つの絵文字につき1000件のポストを用いているのに対し、後者は限られた人数のアノテータによる評価を用いている。さらに、前者のポストの収集時期は2022/8/18~2022/9/23で、後者のアノテーション評価の実施時期は、2024/8/5~2025/12/15であり、前者は後者と比べ約2、3年早く収集されており、その期間に絵文字の使用傾向や社会的な状況が変化している可能性がある。特に、前者のポスト収集期間は約1ヶ月間と短期間であるため、ポストの話題に偏りが生じ、それが結果に影響する可能性がある。

加えて、前者では収集するポスト言語を日本語と英語に限定しているものの、ポスト投稿者の居住地は特定していない。特に、英語は多くの国で使用されている言語であるため、複数の文化圏の使用実態が混在している可能性がある。一方、人手アノテーションは評価人数や年齢層が限定されているため、アノテータ自身の特定の属性がより強く反映されると考えられる。

以上のことから、両手法はそれぞれ異なる側面からも感情を表現しており、そこから抽出された文化差が必ずしも一致しないことは自然な結果であると考えられる。この違いは、絵文字に込められた感情が文脈に依存していること、文化圏や年齢、時期などさまざまな要因で絵文字の使用が異なることを示していると考えられる。

### 7.3 提案手法による文化差抽出結果の妥当性

本研究では、感情予測モデルに基づく絵文字感情ベクトルを用いて文化差のある絵文字を抽出した。その妥当性を検証するため、6.4節では人手ベースの文化差絵文字集合を基準とし、提案手法による抽出結果との一致度を評価した。その結果、提案手法と人手の両方で文化差ありと抽出された絵文字(TP)は137件、提案手法のみで文化差が抽出された絵文字(FP)は168件、人手のみで文化差が抽出された絵文字(FN)は294件であった。このとき適合率は0.45、再現率は0.32、F値は0.37であり、提案手法は一定数の人手でも文化差のある絵文字を抽出できている一方で、人手で文化差が抽出された絵文字を十分に網羅できて

いないことが示された。本節では、文化差抽出結果の妥当性について、文化差の種類別、感情軸別、絵文字カテゴリ別に考察を行う。

### 7.3.1 文化差種類別内訳から見た文化差抽出結果

各文化差において、提案手法では第一種文化差と第二種文化差として抽出された絵文字数が多く、第三種文化差として抽出された絵文字数が少なかった。一方、人手ベースでは、第一種文化差と第二種文化差として抽出された絵文字数が少なく、第三種文化差として抽出された絵文字数が多かった。しかし、第三種文化差において、提案手法での文化差抽出数は少ないものの、適合率が **0.57** と比較的高かった。そこで、提案手法の第三種文化差の抽出数は少ないが、人手ベースの第三種文化差が抽出された絵文字集合と・提案手法における他の二種類の文化差が抽出された絵文字集合と一致、整合しているのではないかと仮説を立て、人手ベースで第三種文化差が抽出された絵文字の内訳を調査した。

表 24 にその内訳を示す。その結果、人手で構造的文化差と判断された絵文字のうち、提案手法では文化差なしと判定されたものが **283** 件と最も多く、提案手法は構造的文化差そのものを直接抽出する感度が十分ではないことがわかる。一方で、提案手法において局所的文化差のみとして抽出されたものが **64** 件、局所的文化差と対極的文化差の両方として抽出されたものが **46** 件存在した。これらは、提案手法が文化差を全体として捉えるというよりも、各感情軸に沿った差

表 24 人手ベースで第三種文化差が抽出された絵文字(416 件)に対する内訳

提案手法で文化差なし	283
提案手法で局所的文化差のみあり	64
提案手法で対極的文化差のみあり	2
提案手法で局所的文化差と対極的文化差あり	46

表 25 人手で第三種文化差ありと判断された絵文字 (416 件) に対する提案手法における低次差分と構造的差分の抽出内訳

	提案手法で 構造的文化差あり	提案手法で 構造的文化差なし
提案手法で 低次の文化差あり	17	112
提案手法で 低次の文化差なし	4	283

分に基づいて文化差を抽出するため、構造的文化差を低次の成分として分解して抽出している可能性がある。提案手法は構造的文化差全体には十分敏感ではないものの、その背景にある個別の感情軸方向の違いには反応していると考えられる。また、人手評価では第三種文化差として抽出された絵文字が多かったことから、特定の感情軸ごとの差異を個別に判断するというよりも、全体的な印象の違いとして総合的に文化差を認識しやすい傾向があると考えられる。

そこで、人手で第三種文化差ありと判断された絵文字 416 件を対象に、提案手法がそれらを第三種文化差として抽出しているのか、低次の文化差として分解して抽出しているのかを検証するため、低次の差分の抽出有無と第三種文化差の抽出有無による 2×2 表を作成した。表 25 に基づき有意水準を 5%( $p \leq 0.05$ ) としてマクネマー検定を行った結果、第三種文化差として抽出される場合よりも、低次差分として抽出される場合が有意に多かった( $p=1.15e-23$ )。提案手法は第三種文化差そのものを直接抽出する感度は低い一方で、第三種文化差を構成する低次の差分である第一種、第二種文化差には高い感度を持つことが示唆された。

表 26 提案手法における第一種文化差での感情軸ごとの TP・FP の内訳

感情	TP の件数	FP の件数
喜び	27	18
悲しみ	32	27
期待	55	60
驚き	55	81
怒り	26	24
恐れ	13	19
嫌悪	22	21
信頼	48	62

表 27 提案手法における第二種文化差での感情軸ごとの TP・FP の内訳

感情	TP の件数	FP の件数
喜びー悲しみ	22	21
期待ー驚き	12	16
怒りー恐れ	12	15
嫌悪ー信頼	42	46

### 7.3.2 感情別内訳から見た文化差抽出結果

提案手法において、第一種文化差、第二種文化差として抽出された絵文字について感情軸ごとに TP・FP に該当する絵文字数を集計し、文化差抽出結果に感情軸で差があるかを分析した。集計の結果を順に表 26, 表 27 に示す。複数の感情で文化差が抽出される場合もあるため、件数は重複を含む。

表 26 より、第一種文化差では、喜び、悲しみ、怒り、嫌悪軸では FP よりも TP に該当する絵文字数が多く、期待、驚き、恐れ、信頼軸では FP に該当する絵文字の方が多かった。特に、驚き軸では TP が 55 件、FP が 81 件であり、提案手法で人手と一致する文化差の抽出ができていたが、過剰な文化差の抽出も多かったと考えられる。

続いて、表 27 より第二種文化差では、喜び—悲しみ軸では FP よりも TP に該当する絵文字数が多く、他の軸では FP に該当する絵文字の方が多かったが、その件数差は第一種文化差に比べ小さかった。嫌悪—信頼軸での文化差抽出数が最も多く、TP が 42 件、FP が 46 件であった。そのため第二種文化差では、どの感情軸でもある程度人手と一致する文化差の抽出ができており、極端に過剰な文化差の抽出が見られる感情軸は見られなかったと考えられる。

### 7.3.3 絵文字カテゴリ別に見た文化差抽出結果

さらに文化差抽出結果の特徴を詳細に把握するために、Unicode の絵文字分類データ<sup>iv</sup>を使用し、絵文字 group ごとに TP・FP・FN を集計し、絵文字グループ別の適合率・再現率・F 値を算出した結果を表 28 に示す。

その結果、「Smileys & Emotion」や「Travel & Places」では F 値が約 0.50 と比較的高く、提案手法と人手の抽出結果が一致しやすい傾向が見られた。これらのカテゴリには感情表現と直接結びつきやすい絵文字が多く、人手でも文化差が明確に認識されやすい領域であると考えられる。一方で、「Food & Drink」では F 値が 0.07 と著しく低く、人手では文化差が抽出されたにもかかわらず、提案手法では十分に抽出できていないことが分かった。このようなカテゴリは生活文化や使用文脈への依存が強く、単純な感情値の差としては捉えにくい可能性がある。また、「People & Body」では再現率が 0.50 と比較的高い一方で、適合率が 0.37 に留まっており、提案手法が文化差の候補を広く拾うものの、人手では感じにくい文化差が一定数含まれていることが考えられる。これは、身体表

---

<sup>iv</sup> <https://unicode.org/Public/emoji/15.1/emoji-test.txt>

現やジェスチャーの解釈が文化圏によって多様であることが影響している可能性がある。

さらに、大分類の中でも一致度が高い・低い領域をより詳細に確認するため、Unicode 絵文字カテゴリのサブグループごとに同様の集計を行った結果を表 29 に示す。なお、サンプル数が極端に少ない分類では指標が不安定になるため、本研究では絵文字数が 10 件以上となるサブグループのみを対象として算出した。その結果、「person-gesture」や「person-activity」では F 値が 0.70 以上となり、人の動作やジェスチャーに関する絵文字は提案手法と人手で文化差が比較

的一致しやすい傾向が確認された。また、「place-building」や「sky & weather」など、場所や自然現象に関する絵文字でも F 値が 0.57~0.69 と比較的高く、文脈に依存しにくいカテゴリでは文化差が安定して抽出される可能性が示唆される。一方で、「food-prepared」や「plant-other」、「light & video」では TP がほとんど得られず、F 値が 0 となった。これらのカテゴリは、人手では文化差が抽出されているにもかかわらず、提案手法では十分に検出できていないことを示している。特に食文化や生活用品に関する絵文字は、使用される場面や文化的背景が多様であるため、感情予測モデルに基づく提案手法では文化差を捉えにくい可能性がある。以上より、提案手法による文化差抽出の妥当性は絵文字の意味カテゴリによって偏りがあり、感情表現や自然現象のように比較的普遍的なカテゴリでは一致度が高い一方で、生活文化的背景や文脈依存性が強いカテゴリでは一致度が低下することが明らかとなった。

表 28 グループ別の文化差抽出精度

グループ	TP	FP	FN	適合率	再現率	F 値
Travel & Places	33	24	42	0.579	0.440	0.500
Smileys& Emotion	24	14	35	0.632	0.407	0.495
People & Body	28	47	28	0.373	0.500	0.427
Activities	10	6	25	0.625	0.286	0.392
Symbols	6	15	8	0.286	0.429	0.343
Animals& Nature	15	18	41	0.455	0.268	0.337
Objects	18	34	66	0.346	0.214	0.265
Food & Drink	2	8	44	0.200	0.043	0.071

表 29 subgroup 別の文化差抽出精度

group	subgroup	TP	FP	FN	適合率	再現率	F 値
People & Body	person-gesture	6	4	0	0.600	1.000	0.750
People & Body	person-activity	6	4	1	0.600	0.857	0.706
Travel & Places	place-building	10	3	6	0.769	0.625	0.690
Travel & Places	sky & weather	8	3	9	0.727	0.471	0.571
Travel & Places	transport-ground	9	10	7	0.474	0.562	0.514
Animals & Nature	animal-mammal	10	9	11	0.526	0.476	0.500
Activities	event	3	1	6	0.750	0.333	0.462
People & Body	person-role	3	3	4	0.500	0.429	0.462
Activities	sport	4	2	9	0.667	0.308	0.421
Smileys & Emotion	face-concerned	4	2	9	0.667	0.308	0.421
People & Body	person-fantasy	3	8	1	0.273	0.750	0.400
Objects	tool	3	4	6	0.429	0.333	0.375
Animals & Nature	animal-bug	2	5	5	0.286	0.286	0.286
Objects	book-paper	2	2	8	0.500	0.200	0.286
People & Body	person-sport	2	7	4	0.222	0.333	0.267
People & Body	body-parts	2	6	6	0.250	0.250	0.250
Animals & Nature	animal-bird	1	2	7	0.333	0.125	0.182
Objects	clothing	1	4	5	0.200	0.167	0.182
Food & Drink	food-asian	1	1	10	0.500	0.091	0.154
Objects	household	1	5	10	0.167	0.091	0.118

Animals & Nature	plant-other	0	2	10	0.000	0.000	0.000
Food & Drink	food-prepared	0	1	12	0.000	0.000	0.000
Objects	light & video	0	2	9	0.000	0.000	0.000

### 7.3.4 分析手法の制約

本研究では、日英間における絵文字の感情表現の違いを抽出するために、順位相関分析、感情対極に基づく  $t$  検定、感情ごとの  $t$  検定という複数の統計的手法を用いた。これにより、異なる観点から文化差を捉えることを試みた一方で、各手法には固有の制約が存在する。

順位相関分析は、感情値の大小関係に着目することで、全体的な感情構造の違いを捉える点に強みがある。しかし、順位のみを扱うため、感情値の数値的な差の大きさは考慮されない。その結果、数値的には小さな差であっても順位が入れ替わることで大きな差として抽出される場合がある一方、数値的には大きな差が存在していても、順位差としては小さく表れる場合がある。また、相関係数が小さい場合には  $p$  値が大きくなりやすく、偶然による差として扱われる可能性がある。さらに、本分析では 8 つの感情を独立したカテゴリとして扱っており、感情間の関連性や共起関係は考慮されていない。

感情対極に基づく  $t$  検定では、対極に位置する感情の差分を用いることで、感情の方向性に着目した分析を行っている。しかし、この手法では、日英それぞれで顕著な差が存在していたとしても、差分を取ることで値が相殺され、結果として 0 に近い値となる場合がある。そのため、実際には差が存在していても、統計的に抽出されない可能性がある点が制約として挙げられる。

感情ごとの  $t$  検定では、感情値の平均差に基づいて日英間の違いを抽出している。しかし、人手アノテーションによる感情ベクトルはサンプル数が限られているのに対し、提案手法による感情ベクトルは大量の投稿データを基に構築されているため、両者の間で統計的検定における  $p$  値の感度に大きな差が生じる。このため、文化差として抽出される絵文字の件数にも手法間で大きな差が生じており、 $p$  値の大小が必ずしも文化差の大きさを直接反映しているとは限らない。

以上のように、本研究で用いた各分析手法は、それぞれ異なる側面から文化差を捉える一方で、抽出できる差の性質や感度には違いがある。したがって、本

研究における文化差の抽出結果は、単一の手法に基づく結論ではなく、複数の分析結果を総合的に解釈する必要がある。

## 7.4 課題

本研究では、感情予測モデルおよび絵文字感情ベクトルを用いて、日英間における絵文字の感情表現の違いを分析した。一方で、本研究の結果を踏まえると、いくつかの今後の課題および改善点が挙げられる。

1 点目はデータに関する課題である。提案手法では、実際の SNS 投稿を用いることで大規模な分析を行っているが、投稿の収集期間が約 1 か月間と限定的であり、特定の話題や社会的状況の影響を受けている可能性がある。今後はより長期間にわたるデータ収集を行うことで、話題の偏りや一時的なトレンドの影響を緩和する必要がある。また、投稿言語を日本語と英語に限定しているものの、投稿者の居住地域や文化的背景までは特定していない。特に世界の公用語である英語のデータは、複数の文化圏の使用実態が混在している可能性がある。地域情報を考慮した分析を行うことで、より詳細な文化差の解釈が可能になると考えられる。

2 点目は感情予測モデルに関する課題である。本研究では、Plutchik の基本 8 感情に基づいて感情を定義し、各感情を独立したカテゴリとして扱っている。しかし、実際の感情表現においては、感情に幅広い強度の差があり、また、感情同士が相互に関連し合うことで感情のニュアンスが変化する場合がある。今後は、感情の強度や感情間の依存関係を考慮した表現手法を導入することで、絵文字に関連する感情をより実世界に近い形で捉えることが可能になると考えられる。また、感情予測モデル自体の精度向上や、他の感情理論に基づく比較も行う必要がある。

3 点目は分析手法に関する課題である。本研究では、順位相関分析および t 検定といった統計的手法を用いて文化差の抽出を行ったが、これらの手法は抽出できる差の性質や感度に限界を持つ。特に、サンプル数の差による p 値の影響や、感情対極に基づく差分処理によって失われる情報があった。さらに、人手での効果量の算出の際に、サンプル数が少ないことが原因で、信頼区間が不安定になることがあった。その結果、既存のライブラリでは、一方の言語で評価値が完全に一致している場合に、効果量が NA になり、文化差の大きさを定量的に評価することが難しい事例もあった。今後は、分析手法の工夫や、人手の評価者数を

増やすことで効果量や信頼区間の推定精度を向上させる必要がある。

以上の課題を改善することで、絵文字を用いた感情表現の文化差についてより実世界に即した文化差を抽出し、詳細な理解が可能になると考えられる。

## 第8章 おわりに

絵文字は、テキストに書き手の感情やイメージを付与する手段として、日常的なコミュニケーションにおいて重要な役割を果たしている。特に、機械翻訳を用いた多言語コミュニケーションにおいては、翻訳によって表現が変化する中で、絵文字が書き手の感情や意図を直接伝える要素として機能する。一方で、絵文字は同一の文字であっても文化によって異なる解釈がなされる場合があり、異言語間で誤解を生じさせる要因となり得る。

本研究では、日本語と英語における絵文字の感情表現に着目し、その文化的な違いを定量的に抽出する手法を提案した。具体的には、X から収集した絵文字付きテキストを感情予測モデルに入力し、8つの感情カテゴリに対する予測値を得ることで、絵文字が抱く感情を8次元のベクトルとして表現した。本研究では、この表現を絵文字感情ベクトルと定義し、日英の絵文字感情ベクトル間の違いから文化差の分析を行った。

本研究では、手法の実現にあたり二つの課題に取り組んだ。第一に、絵文字が付与されたテキストから感情を定量化するため、複数の感情を同時に扱う感情予測モデルを構築した。主観的感情が付与されたデータセットを用いて事前学習済み BERT のファインチューニングを行い、日本語および英語に対応した感情予測モデルを作成した。第二に、絵文字感情の文化差を抽出するため、順位相関分析、各感情軸に基づく t 検定、感情対極に基づく t 検定という複数の統計的手法を用いた分析を行った。さらに、人手による絵文字感情アノテーション結果や実際の投稿文脈の分析を通じて、統計的に抽出された文化差の妥当性を検証した。

本研究の貢献は以下の2点である。

### 絵文字感情モデルの構築

絵文字が付与されたテキストから複数の感情を同時に推定する感情予測モデルを構築した点である。本研究では、8種類の感情カテゴリに基づくマルチラベル感情予測モデルを構築し、日本語および英語のテキストに対して適用可能なモデルを実現した。これにより、絵文字が使用される文脈に含まれる感情を定量的に推定し、分析に用いる基盤を整備した。

### 絵文字感情の文化差抽出手法

絵文字感情ベクトルを用いて、日英間における絵文字の感情表現の文化差を

定量的に抽出する分析枠組みを示した点である。順位相関分析、各感情軸に基づく  $t$  検定、感情対極に基づく  $t$  検定といった複数の統計的手法を用いることで、異なる観点からそれぞれ第一種文化差、第二種文化差、第三種文化差として文化差の抽出を行った。さらに、人手評価との比較を通じて、統計的に抽出された文化差の妥当性を確認した結果、適合率は **0.45**、再現率は **0.32**、 $F$  値は **0.37** であった。また、人手評価の文化差と比較して、第三種文化差がより低次の差分として第一種文化差や第二種文化差として抽出される傾向にあることもマクネマー検定によって示唆された。

今後の課題としては、より長期間かつ大規模な投稿データを用いた分析や、投稿者の地域情報を考慮した文化差分析が挙げられる。また、感情間の依存関係や共起構造を考慮したモデルおよび分析手法を導入することで、絵文字に付随する感情表現をより精緻に捉えることが可能になると考えられる。本研究で提案した手法は、絵文字を用いた感情表現の文化差理解を深める一助となるものであり、今後の多言語コミュニケーション研究や感情分析研究への応用が期待される。

## 謝辞

本研究を行うにあたり，ご指導いただきました村上陽平教授，Mondheera Pituxcoosuvam 講師に深く感謝申し上げます。

また，アノテーション作業にご協力いただきました皆様に感謝申し上げます。

## 参考文献

- [1] Holtgraves, T., Robinson, C.: Emoji can facilitate recognition of conveyed indirect meaning. In: PLOS ONE, 15(4), e0232361 (2020).
- [2] Kralj Novak, P., Smailović, J., Sluban, B., Mozetič, I.: Sentiment of Emojis. In: PLOS ONE, 10(12), e0144296 (2015).
- [3] Zhao, P., Jia, J., An, Y., Liang, J., Xie, L., Luo, J.: Analyzing and Predicting Emoji Usages in Social Media. In: Companion Proceedings of the The Web Conference 2018, pp.327–334 (2018).
- [4] Brants, W., Sharif, B., Serebrenik, A.: Assessing the Meaning of Emojis for Emotional Awareness - A Pilot Study. In: Companion Proceedings of The 2019 World Wide Web Conference, pp.419–423 (2019).
- [5] Leonardi, V.: Communication challenges and transformations in the Digital Era: emoji language and emoji translation. In: Language and Semiotic Studies, 8(3) (2022).
- [6] Cho, H., Ishida, T., Yamashita, N., Inaba, R., Mori, Y., Koda, T.: Culturally-Situated Pictogram Retrieval. In: International Workshop on Intercultural Collaboration, Lecture Notes in Computer Science, vol.4568, pp.221–235 (2007).
- [7] Dresner, E., Herring, S. C.: Functions of the Nonverbal in CMC: Emoticons and Illocutionary Force. In: Communication Theory, 20(3), pp.249–268 (2010).
- [8] Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp.4171–4186 (2019).
- [9] Kajiwara, T., Chu, C., Takemura, N., Nakashima, Y., Nagahara, H.: WRIME: A New Dataset for Emotional Intensity Estimation with Subjective and Objective Annotations. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp.2095–2104 (2021).

- [10]Plutchik, R.: A General Psychoevolutionary Theory of Emotion. In: Theories of Emotion, pp.3–33 (1980).
- [11]Cohen, J.: Statistical Power Analysis for the Behavioral Sciences. 2nd ed., Routledge, New York (1988).
- [12]Byrt, T., Bishop, J., Carlin, J. B.: Bias, prevalence and kappa. In: Journal of Clinical Epidemiology, 46(5), pp.423–429 (1993).
- [13]Cliff, N.: Dominance statistics: Ordinal analyses to answer ordinal questions. In: Psychological Bulletin, 114(3), pp.494–509 (1993).
- [14]Torchiano, M.: effsize: Efficient Effect Size Computation. R package version 0.8.1, CRAN (2020).

# 付録

## A.1 提案手法と人手評価における TP・FP・FN の絵文字一覧

<p>TP(137 件)</p>	
<p>FP(168 件)</p>	
<p>FN(294 件)</p>	

